

UNIVERSIDAD POLITÉCNICA DE MADRID



POLITÉCNICA

**ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE
TELECOMUNICACIÓN**



Tesis Doctoral

**Modelado de sistemas multimedia para
personalización y recomendación híbrida a
partir del consumo audiovisual de los
usuarios**

FAUSTINO ÁNGEL SÁNCHEZ GARCÍA
Ingeniero de Telecomunicación

2012

SIGNALS, SYSTEMS AND RADIOCOMMUNICATIONS

DEPARTMENT

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE TELECOMUNICACIÓN

DOCTORAL THESIS

**Modelling of multimedia systems for
personalization and hybrid recommendation
based on users audiovisual consumption**

Author:

Faustino Ángel Sánchez García

Ingeniero de Telecomunicación

Directors:

Federico Álvarez García

Doctor and Ingeniero de Telecomunicación

Guillermo Cisneros Pérez

Doctor and Ingeniero de Telecomunicación

December 2012

DEPARTAMENTO DE SEÑALES, SISTEMAS Y
RADIOCOMUNICACIONES

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE TELECOMUNICACIÓN

TESIS DOCTORAL

**Modelado de sistemas multimedia para
personalización y recomendación híbrida a partir
del consumo audiovisual de los usuarios**

Autor:

Faustino Ángel Sánchez García

Ingeniero de Telecomunicación

Directores:

Federico Álvarez García

Doctor e Ingeniero de Telecomunicación

Guillermo Cisneros Pérez

Doctor e Ingeniero de Telecomunicación

Diciembre 2012

UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE
TELECOMUNICACIÓN

DEPARTAMENTO DE SEÑALES, SISTEMAS Y RADIOCOMUNICACIONES

“Modelado de sistemas multimedia para personalización y recomendación híbrida a partir del consumo audiovisual de los usuarios”

Autor: Faustino Ángel Sánchez García

Directores:

Dr. Federico Álvarez García

Dr. Guillermo Cisneros Pérez

El tribunal nombrado para juzgar la tesis arriba indicada el día de de 2012,
compuesto de los siguientes doctores:

Presidente: _____

Vocal: _____

Vocal: _____

Vocal: _____

Secretario: _____

Realizado el acto de lectura y defensa de la Tesis Doctoral, en Madrid, a de
de 2012, acuerdan otorgarle la calificación de:

Calificación _____

EL PRESIDENTE

LOS VOCALES

EL SECRETARIO

FINAL DE PARTIDA

Recopilar para avanzar. Hacer balance como medio de subsistencia. Fabricar listas como quien dibuja su alma. Soñar a Perek como quien baila con fantasmas. Evocar a Deleuze, la imagen rizomática, el recuerdo inaprensible, el pasado que se fuga por la chimenea. Esculpir el tiempo y dibujar su evanescencia. Ser postmodernos y referenciales. Pensar, como Idea Vilariño, en lo que ya no será, porque esa es a veces la mejor forma de reconocer lo que fue.

Ya no será tiempo de tesis, siempre ligado al Gato, que no es más que quienes formamos y han formado parte de él. De ellos son inseparables estos casi cinco años de doctorado, casi seis desde que el grupo me abrió las puertas y fui conociendo a los que estaban, a los que llegaron conmigo y a las sucesivas generaciones que he ido viendo pasar durante estos años. Este tiempo de tesis no habría sido posible sin ellos, desde José Manuel hasta todos los doctorandos y proyectandos del grupo: Juan, Jorge, Marcos, Ana, Alberto, Álex Cuervo, Carlos Alberto, Adrián, Diego, Iago, Silvia, Amaya, Vero, David, Ana, Juanpe, y todos los demás. Especialmente importantes han sido los chicos que han trabajado conmigo, desde José Manuel, mi primer proyectando, hasta Paula, Orlando, Agustín, Borja o Jon, a quienes también pertenece un pedazo importante de esta tesis. Y sin olvidar a Marta, que ha sido mucho más que compañera de mesa y de travesías en proyectos y papers durante la segunda mitad de la tesis.

Ya no será tiempo de directores, sin los cuales esta tesis no habría sido posible: Guillermo, siempre dispuesto cuando es necesario a pesar de la dictadura de las agendas y, especialmente, Fede, gracias a quien he podido llevar una línea de investigación tan precisa y coherente, gracias a quien esos proyectos en los que he trabajado eran los que tenían que ser, y gracias a quien he tenido el apoyo y la confianza necesarias a la vez que ambos íbamos creciendo y evolucionando.

Ya no será tiempo de tesis, que habría sido un tiempo muerto de no haber tenido, también, sus convulsiones, conflictos y momentos de fricción. No hay peor película que la película perfecta, como podrían decir Godard o Rivette.

Ya no será tiempo de tesis, que también ha sido tiempo de transfusiones entre el trabajo y la vida. La transfusión de Lara, entre el iLab, los cursos asturianos y la intensidad de los momentos que compartimos juntos. La transfusión de Marta (con quien arranqué todo esto en aquel ya lejano 3SME), Sergio y Álex, en aquellas explosiones veraniegas que sacudieron el tiempo del exilio en el C, la de Nuria, entre cenas íntimas, quimeras, fiestas extrañas y reconfortantes conversaciones infinitas, y la transfusión de Laura, mi compañera de siempre, incluso cuando dejó de sentarse a mi lado, con quien he recorrido todo el camino de la tesis atravesando meandros y circunvoluciones sin soltarnos de la mano. Y todo esto sin olvidar la transfusión de Cris, entre teatros y autobuses y, sobre todo, la transfusión de María, densa pero poderosa como un tsunami, que me trajo tantas cosas, de la Escuela a Lavapiés, de Sol hasta el centro de nuestras conciencias, deseos, alianzas, inquietudes; cosas que se ramifican y enlazan con otras, de Rodrigo al párrafo siguiente, desde nuestra voluntad psicoanalista hasta la hora en que Buscamedia desaparece para convertirse en el más poderoso de los cimientos.

Ya no serán muchas cosas que han sido fuera de la Escuela, aunque también viven de ciertas transfusiones del pasado y del presente. Transfusiones que quizá empezaron a

forjarse hace una década, en algún laboratorio de la carrera, cuando comenzó con Silvia un bucle que no tendría fin, y que con ella sería eterno. Y transfusiones que seguirían con César también por aquellos años, cuando éramos más jóvenes y no sabíamos que acabaríamos los dos haciendo en la Escuela un doctorado que nos haría pensar en Montescos y Capuletos mientras chateábamos sobre películas que nadie quería ver con nosotros. Y ahí apareció Jon, que me abrió una puerta impagable a un mundo que era más cercano al mío, el de las películas y exposiciones, discusiones de libros y filmoteca, que, de repente, dejaba de ser mi lugar habitual de peregrinaje solitario para convertirse en un nuevo espacio social. Con Jon llegó Antonio y las infinitas cadenas de mails no sólo cinéfilos, y Javi, mucho más allá de HTML5, somnolencias y papers, y Elena, tan especial en Madrid como en París, aunque ya no puedan ser tan habituales los paseos nocturnos y los colacaos. Y por último, hace apenas un año, esa irrupción fulgurante, luminosa, en pleno festival de San Sebastián, de Alicia y Gonzalo, como un *Tourbillon de la vie*, para completar de cuadros, collages, libros y entusiasmos cinéfilos ese grupo que siempre está abierto y que ha permitido que esta última época haya sido la mejor.

Ya no será ese tiempo soñado, que como el tiempo filmado o el tiempo narrado, son los únicos que permiten vivir dos veces, y por eso estos años son inseparables de mi “otra vida”, la del cine, la literatura, las revistas... Desde esos primeros relatos publicados por Eloy en El problema de Yorick, hasta los devaneos en los festivales de cine de la mano de Mateo o las incursiones radiofónicas con Ayoze, pasando por los editores de Detour, El rayo verde, Transit y, sobre todo, de Shangrila, Jesús Rodrigo, un ejemplo por su seriedad y rigor y que, además, me permitió descubrir a Aarón Rodríguez y la Cinefilia 2.0. Porque parte de esta tesis nace de las influencias multidisciplinares, de años de interiorizar resortes del cine, la imagen y el misterio. Y la palabra, siempre la palabra.

Ya no serán los que serán siempre, mis padres y mi hermana, a quienes debo todo aunque sea lo más difícil expresar, entre el pudor y lo sobreentendido, porque son el sustento, la base de mí por dentro y por fuera, la explicación de tanto Proust y Dostoievski, e incluso de los Mizoguchi de los que huye mi padre o los Godard que asustan a mi madre. Ya no será que mi padre vuelva a preguntarme por la tesis aunque me siga preguntando por todas las demás cosas mientras intenta no aparentarlo, como buen investigador. Ya no serán las películas que no podía ver en verano con mi madre y con Inma porque tenía que escribir la tesis. Tampoco serán quienes ya no están aquí, aunque lo hubieran dado todo por ello, como mi abuelos Evaristo (de quien tanto conservo invisible, fantasmal), José y Consuelo, o mi tía Reme, a quien se la llevaron los disgustos y las cosas que nunca querremos saber; mientras tanto, mi abuela María y mi tía Agus resisten los golpes y viven con inmortal ilusión.

Y, en resumen, ya no serán los efluvios evanescentes, los instantes que existieron en el sueño, los recuerdos inventados, el salón en el ángulo oscuro, el dormitorio espectral donde Maud se quita su disfraz garreliano mientras juega a ser Bresson. Ya no serán los viajes imaginarios, como son todos los viajes, la iluminación imposible o las lealtades secretas. Ya no será lo irreplicable. Porque el paso a la madurez es darse cuenta de que lo más importante es aquello que jamás ha sucedido, aquello que la vida juega en sus elipsis imposibles, en sus descabelladas ironías. El relato soñado, la imagen evocada, el efluvio otra vez, el efluvio despierto. La vida es sueño, susurrando más cerca de Schnitzler que de Calderón.

Ya no serán porque, afortunadamente, seguirán siendo otros. Otros irreplicables.

Índice de contenidos

1.	Introducción.....	3
1.1	Contexto de los Sistemas de Recomendación	3
1.2	Retos actuales de los sistemas de recomendación	4
1.3	Aportes de la tesis.....	6
1.4	Organización de la tesis	8
2.	Objetivos y metodología.....	9
2.1	Creación de un modelo que permita inferir preferencias de los usuarios a partir de su consumo audiovisual	10
2.1.1	Descripción.....	10
2.1.2	Metodología.....	10
2.2	Creación de un modelo que permita describir imágenes y vídeos en función de factores influyentes en los gustos de los usuarios e independientemente de su contenido semántico	11
2.2.1	Descripción.....	11
2.2.2	Metodología.....	12
2.3	Creación de un sistema de recomendación personalizada, cuyos criterios se adapten dinámicamente a las preferencias de cada usuario.....	13
2.3.1	Descripción.....	13
2.3.2	Metodología.....	14
3.	Técnicas y arquitecturas de sistemas de recomendación multimedia.....	15
3.1	Modelado general y requisitos de los escenarios.....	15
3.2	Adaptación a distintos escenarios.....	18
3.2.1	Arquitectura con recomendación en cliente	18
3.2.2	Arquitectura con recomendación en cliente o baja carga en servidor	20
3.2.3	Arquitectura con recomendación en servidor.....	22
4.	Modelado del consumo audiovisual del usuario para la obtención del interés	27
4.1	Introducción.....	27
4.2	Estado del arte	28
4.3	Captura de datos de entrada de fuentes heterogéneas (información explícita e implícita).....	31
4.4	Modelo de transformación de consumo en interés	33
4.4.1	Medida y cuantificación del consumo audiovisual (interés local)	36
4.4.2	Cálculo del interés acumulado del usuario (interés global).....	40
4.5	Aplicación a entornos categorizados	50
4.6	Aplicación del algoritmo a entornos multiatributo.....	55
4.7	Pruebas y validación.....	62
4.7.1	Pruebas en entornos categorizados	62
4.7.2	Pruebas en entornos multiatributo	71
4.8	Resultados obtenidos	80
5.	Análisis y clasificación de contenidos audiovisuales para la recomendación.....	81
5.1	Introducción.....	81

5.2	Estado del arte	82
5.3	Modelo de extracción de características y anotación de imágenes para la recomendación.....	85
5.4	Modelo de extracción de características y anotación de vídeos para la recomendación.....	90
5.4.1	Caracterización de atributos fílmicos	90
5.4.2	Modelo de descriptores fílmicos para la recomendación	91
5.4.3	Modelo de extracción automática.....	95
5.4.4	Capa de metadatos	95
5.4.5	Capa visual. Descriptores de bajo nivel.....	103
5.4.6	Capa visual. Descriptores de alto nivel	105
5.5	Pruebas y validación	110
5.5.1	Validación de descriptores de imagen	110
5.5.2	Validación de descriptores de vídeo	112
5.6	Resultados obtenidos	124
6.	Algoritmos de recomendación híbrida	125
6.1	Introducción.....	125
6.2	Estado del arte	127
6.2.1	Recomendación basada en el comportamiento social	127
6.2.2	Recomendación basada en contenido	134
6.2.3	Algoritmos de hibridación	143
6.3	Filtrado colaborativo	148
6.4	Recomendación de imágenes basada en contenido	150
6.5	Recomendación de vídeos basada en contenido.....	159
6.6	Explicación	162
6.7	Hibridación, pruebas y validación	163
6.7.1	Recomendador de imágenes	163
6.7.2	Recomendador de vídeos.....	171
6.8	Resultados obtenidos	176
7.	Conclusiones y líneas futuras	177
7.1	Conclusiones.....	177
7.2	Avance del conocimiento y divulgación	181
7.3	Líneas futuras	183
8.	Referencias	185
	Acrónimos	199
	ANEXO I: MODELO DE ENCUESTA PARA LAS PRUEBAS DE MODELADO DE CONSUMO	201
	ANEXO II: ALGORITMOS EMPLEADOS PARA LA OBTENCIÓN DE LOS DESCRIPTORES DE BAJO NIVEL.....	203
	ANEXO III: CORPUS SELECCIONADO PARA EL RECOMENDADOR DE IMÁGENES	207
	ANEXO IV: CORPUS SELECCIONADO PARA EL RECOMENDADOR DE VÍDEOS	213

Índice de figuras

Fig. 1. Escenario general de distribución y recomendación de contenidos audiovisuales	9
Fig. 2. Tipos de metadatos del objeto multimedia [10]	16
Fig. 3. Arquitectura en un escenario <i>broadcast</i>	19
Fig. 4. Arquitectura en un escenario concurrente	21
Fig. 5. Arquitectura orientada a búsqueda	22
Fig. 6. Conversor consumo-interés. Modelo general	35
Fig. 7. Interés local. 30%. (a)	37
Fig. 8. Interés local. 60%. (b)	37
Fig. 9. Interés local. 60%. (c)	38
Fig. 10. Interés local. 30%. (d)	38
Fig. 11. Cuantificador uniforme empleado	40
Fig. 12. Relación de la conversión del consumo en interés con un HMM	41
Fig. 13. Adaptación de un HMM al problema de inferencia del interés	42
Fig. 14. Relación de la matriz B con sus submatrices	45
Fig. 15. Proceso y ejemplo de la variación de la matriz B	47
Fig. 16. Algoritmo de obtención del interés global	49
Fig. 17. Conversor consumo-interés. Modelo <i>broadcast</i>	52
Fig. 18. Evolución del factor de compañía	54
Fig. 19. Aplicación del modelo general a un entorno multiatributo deportivo	57
Fig. 20. Dependencia de la interrupción del consumo en el interés local	58
Fig. 21. Modelado del atributo “Ronda”	59
Fig. 22. Implementación en un Set-Top box	62
Fig. 23. Identificación del usuario	65
Fig. 24. Correlaciones de Spearman y Pearson para géneros y programas periódicos	68
Fig. 25. Infertaz del sistema de recomendación: usuario 1	75
Fig. 26. Solución alternativa mediante árbol de decisión	76
Fig. 27. Valores de correlación por usuario (umbrales de [31])	78
Fig. 28. Esquema general de recomendación de vídeos	81
Fig. 29. Interfaz avanzada de Cinemetrics	83
Fig. 30. Descriptores seleccionados para la clasificación de las imágenes y ejemplo de HDL	85
Fig. 31. Modelo de descripción formal y estética de vídeos	96
Fig. 32. Detección del movimiento de cámara y del movimiento interno de cuadro	106
Fig. 33. Base de datos necesaria para la validación de los descriptores de vídeo	115
Fig. 34. Histogramas de los valores de los 26 descriptores de vídeo	121
Fig. 35. Valores medios del estadístico Kolmogorov-Smirnov para los usuarios de la base de datos	122
Fig. 36. Porcentaje de usuarios influidos por los descriptores del modelo	122
Fig. 37. Arquitectura de un sistema de recomendación para redes sociales [33]	133
Fig. 38. Modelo de recomendación para fuentes heterogéneas en redes sociales [95]	134
Fig. 39. Formato de tablas de la base de datos	135
Fig. 40. Relaciones entre ítems, contenidos y usuarios	135
Fig. 41. Algoritmo de Rocchio	136

Fig. 42. Niveles de clasificación de vídeos	141
Fig. 43. Cluster de vecinos del usuario A.....	149
Fig. 44. Ejemplo de predicción del rating del ítem B para el usuario A	150
Fig. 45. Esquema de bloques del sistema de recomendación de imágenes	151
Fig. 46. Algoritmo de recomendación basado en perfil	154
Fig. 47. Proceso de recomendación de vídeos basado en contenido	160
Fig. 48. Valor medio de mejora.....	165
Fig. 49. Comparación con la mejora maxima teórica para NR=5	166
Fig. 50. Comparación entre las técnicas empleadas y la máxima mejora teórica para NR=5	170
Fig. 51. Comparación de histogramas según las cuatro técnicas de recomendación ...	174

Índice de tablas

Tabla 1. Características de los escenarios de aplicación	18
Tabla 2. Aplicabilidad de técnicas a los distintos tipos de arquitectura	24
Tabla 3. Descripción de los estados [17]	30
Tabla 4. Ejemplo del sistema de Zimmerman [18]	31
Tabla 5. Resultados de la aplicación del interés local	39
Tabla 6. Caracterización de las distribuciones matriciales	46
Tabla 7. Ejemplo de conversión de consumo a interés (Usuario 9): “Noticias TVE 1”	66
Tabla 8. Correlación entre preferencias implícitas y explícitas (Usuario 9)	67
Tabla 9. Resultados de correlación (10 géneros / 40 programas)	69
Tabla 10. Resultados de correlación sin considerar programas del corazón (9 géneros / 36 programas)	71
Tabla 11. Corpus de recomendación	73
Tabla 12. Simulación de consumos de usuarios	74
Tabla 13. Valores de AR (Importancia de atributo) inferidos por el sistema	74
Tabla 14. Resultados de recomendación del sistema propio	75
Tabla 15. Resultados de recomendación del árbol de decisión	76
Tabla 16. Valores obtenidos y correlaciones	77
Tabla 17. Comparación con otras propuestas	79
Tabla 18. Descriptores incluidos en el modelo	97
Tabla 19. Test de validación de descriptores para la clasificación	112
Tabla 20. Número de películas según el número de valoraciones	113
Tabla 21. Ejemplo de valoraciones del usuario u	118
Tabla 22. Valores de los parámetros de entrada seleccionados para las pruebas	164
Tabla 23. Comparación de técnicas de recomendación de imágenes para NR=5	165
Tabla 24. Ejemplo recomendación basada en contenido para usuario “a”	167
Tabla 25. Ejemplo inicialización recomendación para usuario “u”	168
Tabla 26. Comparativa general de las técnicas de recomendación de imágenes para NR=5	170
Tabla 27. RMSE de las distintas técnicas de recomendación de imágenes	171
Tabla 28. Comparativa general de las técnicas de recomendación de vídeos para NR=10	173
Tabla 29. Comparativa general de las técnicas de recomendación de vídeos para NR=5	174
Tabla 30. Porcentaje de usuarios según umbrales de mejora	175
Tabla 31. Ejemplo de recomendación de vídeo basada en contenido	175
Tabla 32. Ejemplo del algoritmo de inicialización del recomendador de vídeo	175
Tabla 33. Difusión de las investigaciones centrales de la tesis doctoral	181

Abstract

This doctoral thesis focuses on the modeling of multimedia systems to create personalized recommendation services based on the analysis of users' audiovisual consumption. Research is focused on the characterization of both users' audiovisual consumption and content, specifically images and video. This double characterization converges into a hybrid recommendation algorithm, adapted to different application scenarios covering different specificities and constraints.

Hybrid recommendation systems use both content and user information as input data, applying the knowledge from the analysis of these data as the initial step to feed the algorithms in order to generate personalized recommendations.

Regarding the user information, this doctoral thesis focuses on the analysis of audiovisual consumption to infer implicitly acquired preferences. The inference process is based on a new probabilistic model proposed in the text. This model takes into account qualitative and quantitative consumption factors on the one hand, and external factors such as zapping factor or company factor on the other.

As for content information, this research focuses on the modeling of descriptors and aesthetic characteristics, which influence the user and are thus useful for the recommendation system. Similarly, the automatic extraction of these descriptors from the audiovisual piece without excessive computational cost has been considered a priority, in order to ensure applicability to different real scenarios.

Finally, a new content-based recommendation algorithm has been created from the previously acquired information, i.e. user preferences and content descriptors. This algorithm has been hybridized with a collaborative filtering algorithm obtained from the current state of the art, so as to compare the efficiency of this hybrid recommender with the individual techniques of recommendation (different hybridization techniques of the state of the art have been studied for suitability). The content-based recommendation focuses on the influence of the aesthetic characteristics on the users. The heterogeneity of the possible users of these kinds of systems calls for the use of different criteria and attributes to create effective recommendations. Therefore, the proposed algorithm is adaptable to different perceptions producing a dynamic representation of preferences to obtain personalized recommendations for each user of the system.

The hypotheses of this doctoral thesis have been validated by conducting a set of tests with real users, or by querying a database containing user preferences - available to the scientific community. This thesis is structured based on the different research and validation methodologies of the techniques involved. In the three central chapters the state of the art is studied and the developed algorithms and models are validated via self-designed tests. It should be noted that some of these tests are incremental and confirm the validation of previously discussed techniques.

Resumen

Esta tesis doctoral se centra en el modelado de sistemas multimedia para la creación de servicios personalizados de recomendación a partir del análisis de la actividad de consumo audiovisual de los usuarios. La investigación se focaliza en la caracterización tanto del consumo audiovisual del usuario como de la naturaleza de los contenidos, concretamente imágenes y vídeos. Esta doble caracterización de usuarios y contenidos confluye en un algoritmo de recomendación híbrido que se adapta a distintos escenarios de aplicación, cada uno de ellos con distintas peculiaridades y restricciones.

Todo sistema de recomendación híbrido toma como datos de partida tanto información del usuario como del contenido, y utiliza este conocimiento como entrada para algoritmos que permiten generar recomendaciones personalizadas.

Por la parte de la información del usuario, la tesis se centra en el análisis del consumo audiovisual para inferir preferencias que, por lo tanto, se adquieren de manera implícita. Para ello, se ha propuesto un nuevo modelo probabilístico que tiene en cuenta factores de consumo tanto cuantitativos como cualitativos, así como otros factores de contorno, como el factor de *zapping* o el factor de compañía, que condicionan la incertidumbre de la inferencia.

En cuanto a la información del contenido, la investigación se ha centrado en la definición de descriptores de carácter estético y morfológico que resultan influyentes en el usuario y que, por lo tanto, son útiles para la recomendación. Del mismo modo, se ha considerado una prioridad que estos descriptores se puedan extraer automáticamente de un contenido sin exigir grandes requisitos computacionales y, de tal forma que se garantice la posibilidad de aplicación a escenarios reales de diverso tipo.

Por último, explotando la información de preferencias del usuario y de descripción de los contenidos ya obtenida, se ha creado un nuevo algoritmo de recomendación basado en contenido. Este algoritmo se cruza con un algoritmo de filtrado colaborativo de referencia en el estado del arte, de tal manera que se compara la eficiencia de este recomendador híbrido (donde se ha investigado la idoneidad de las diferentes técnicas de hibridación del estado del arte) con cada una de las técnicas individuales de recomendación. El algoritmo de recomendación basado en contenido que se ha creado se centra en las posibilidades de la influencia de factores estéticos en los usuarios, teniendo en cuenta que la heterogeneidad del conjunto de usuarios provoca que los criterios y atributos que condicionan las preferencias de cada individuo sean diferentes. Por lo tanto, el algoritmo se adapta a las diferentes percepciones y articula una metodología dinámica de representación de las preferencias que permite obtener recomendaciones personalizadas, únicas para cada usuario del sistema.

Todas las hipótesis de la tesis han sido debidamente validadas mediante la realización de pruebas con usuarios reales o con bases de datos de preferencias de usuarios que están a disposición de la comunidad científica. La diferente metodología de investigación y validación de cada una de las técnicas abordadas condiciona la estructura de la tesis, de tal manera que los tres capítulos centrales se estructuran sobre su propio estudio del estado del arte y los algoritmos y modelos desarrollados se validan mediante pruebas autónomas, sin impedir que, en algún caso, las pruebas sean incrementales y ratifiquen la validación de técnicas expuestas anteriormente.

1. Introducción

1.1 Contexto de los Sistemas de Recomendación

Los avances tecnológicos de la última década, desde la digitalización de los contenidos y los medios de producción y distribución hasta la popularización de la banda ancha, así como el aluvión de pantallas y dispositivos de acceso a Internet en la vida cotidiana, han permitido que aumente espectacularmente la oferta de contenidos digitales que están a disposición del consumo de los usuarios. Este nuevo marco de consumo ha intensificado viejos problemas relacionados con el marketing o la sociología, al mismo tiempo que facilita la implantación de nuevas técnicas y herramientas para solucionarlo. Estos problemas afectan tanto a los productores y a los vendedores como a los usuarios. Por un lado, los productores tienen la necesidad de encontrar el público adecuado al que dirigir su oferta comercial; por otro, los usuarios se ven desbordados ante el exceso de oferta, en ocasiones porque tampoco son capaces de determinar cómo es aquello que quieren o que les gusta, o simplemente desean descubrir productos hasta el momento desconocidos para ellos dentro de un determinado dominio de conocimiento.

Hasta el momento de la digitalización, las estrategias para combatir estos problemas eran limitadas. Desde el punto de vista del vendedor, la manera de encontrar un público adecuado era la realización de estudios de mercado que fueran capaces de segmentar a los usuarios en grupos sobre los que extraer determinadas necesidades o preferencias. En cualquier caso, además de muy costoso, resultaba inabarcable realizar estudios individualizados, que permitieran el desarrollo de estrategias publicitarias personalizadas. Por lo tanto, los avances en técnicas y sistemas para el tratamiento de la información en las últimas décadas cambiaron el paradigma, de tal manera que, especialmente en el caso del comercio electrónico, los vendedores empezaron a verse en condiciones de ofrecer a cada usuario aquellos productos que tenían más probabilidad de vender, aquellos que satisficieran en mayor medida sus necesidades.

En este sentido, una de las estrategias más populares de los últimos años ha sido la implantación de sistemas de recomendación personalizada, de tal manera que un sistema, automáticamente, recomendara a cada usuario aquello que hipotéticamente, de acuerdo con el conocimiento adquirido sobre él, más pudiera atraerle. Esta es la razón por la que la mayor parte de los portales de Internet orientados a la venta de productos, tanto físicos como digitales, hayan incorporado (o lo estén haciendo) sistemas de recomendación adecuados a la naturaleza de sus productos.

Desde el punto de vista del usuario, sin embargo, el problema tradicional era doble: por un lado, encontrar aquello que éste buscaba cuando sabía cómo era lo que estaba buscando y, por otro, encontrar aquello que no sabía cómo era. En ambos casos, la solución era la misma, preguntar o dejarse aconsejar por expertos o por personas de confianza que dispusieran de un mayor conocimiento sobre un determinado tema. Es decir, el usuario buscaba recomendadores personales que pudieran servir de orientación ante la imposibilidad de disponer uno mismo de todo el conocimiento. Estos

recomendadores podían ser públicos, como los críticos de cine o los críticos literarios, que realizaban recomendaciones generales para un gran número de usuarios (es en este caso el propio usuario quien debe adquirir conocimiento sobre el crítico para poder catalogarlo entre sus recomendadores de confianza), o privados, como el caso de personas del entorno cuya opinión resulte fiable en un determinado ámbito. El nuevo paradigma digital no ha eliminado estas formas de recomendación, pero ha añadido la posibilidad de contar con estrategias más potentes, todavía en plena evolución, que permitan tener en cuenta la totalidad del conocimiento que se dispone sobre los usuarios y sobre los contenidos y generar así críticos personalizados para cada usuario. Además, si en viejo paradigma el éxito de un crítico radicaba en ganarse la confianza de un sector de usuarios, aquel que le fuera más afín, en el nuevo paradigma, un sistema de recomendación exitoso debe poder ganarse la confianza de usuarios heterogéneos, ya que debe tener la capacidad de adaptarse dinámicamente al gusto de estos. Este es el papel de los sistemas de recomendación.

Por estos motivos han irrumpido con tanta fuerza en los últimos años las investigaciones alrededor de los sistemas de recomendación, aplicadas a los más diversos ámbitos y centrados en la recomendación tanto de ítems genéricos como de productos específicos o de una determinada naturaleza. A nivel comercial, esto queda patente comprobando la gran cantidad de sistemas de recomendación que se han implantado en portales de Internet de diferente orientación, ya sea a través de la recomendación de productos, como Amazon¹ o ChoiceStream², de noticias, como Daily Me³ o Reddit⁴, de música, como Last.fm⁵ o Pandora⁶, o de películas, como Filmaffinity⁷ o Netflix⁸. Dentro del campo de los sistemas de recomendación, por lo tanto, se aprecia un interés especial en el dominio de los contenidos audiovisuales, a causa de las ventajas en la producción y distribución contenidos que ha acarreado la digitalización de los recursos multimedia.

En el caso de esta tesis, las aportaciones contribuyen a la generación de conocimiento relativa a la recomendación de contenidos multimedia y, más concretamente, a la recomendación de imágenes y vídeos.

1.2 Retos actuales de los sistemas de recomendación

La juventud de la disciplina de los sistemas de recomendación provoca, a pesar del creciente interés de la comunidad científica internacional, que existan todavía numerosos asuntos sin abordar, especialmente dentro de la recomendación de contenidos multimedia. Estos asuntos se irán desgranando en los sucesivos capítulos de la tesis, al mismo tiempo que se presentarán las nuevas soluciones generadas, pero, a grandes rasgos, los retos del estado del arte referidos a los sistemas de recomendación se pueden agrupar en cuatro grandes bloques, derivados de la idea de que una recomendación se basa en la conjunción del conocimiento que se dispone tanto del producto susceptible de ser recomendado como del usuario a quien se va a recomendar. Por lo tanto, la posibilidad de éxito de una recomendación dependerá tanto de recabar la

¹ www.amazon.com (consultada el 20/09/2012)

² www.choicestream.com/ (consultada el 20/09/2012)

³ www.dailyme.com/ (consultada el 20/09/2012)

⁴ www.es.reddit.com/ (consultada el 20/09/2012)

⁵ www.lastfm.es/ (consultada el 20/09/2012)

⁶ www.pandora.com/ (consultada el 20/09/2012)

⁷ www.filmaffinity.com (consultada el 20/09/2012)

⁸ www.netflix.com (consultada el 20/09/2012)

información apropiada y gestionarla de manera que aporte el mayor conocimiento posible respecto a las preferencias interiorizadas de cada persona, como de utilizar los algoritmos y métricas más adecuados para generar predicciones que permitan extrapolar de la manera más precisa posible el conocimiento de las preferencias del usuario sobre unos ítems a otros ítems cuya naturaleza ha sido previamente caracterizada:

1. Adquisición y modelado de la información del usuario

Es necesario investigar fórmulas nuevas y flexibles que posibiliten adquirir la mayor información posible (y útil de cara a la recomendación) de los usuarios sin resultar invasivo y de la manera más transparente posible. Para ello, es necesario tener en cuenta el tipo de plataforma en el que se desarrolla el sistema de recomendación, y que sirve como punto de acceso a la información del usuario. Ya dentro del caso de recomendación multimedia, resulta especialmente interesante considerar plataformas de distribución de contenidos, donde se plantea como un reto fundamental caracterizar las preferencias de los usuarios a partir del consumo audiovisual, ya que esta actividad no presenta ningún añadido a la actividad habitual del usuario dentro de la plataforma. Otra importante línea de investigación consiste en el modelado de información heterogénea el usuario, procedente de fuentes diversas y de distinta naturaleza [1], ya sea consumo, opiniones explícitas sobre un ítem expresadas mediante una valoración, acciones ejecutadas sobre el ítem (ya sea compartir, incluir en favoritos, etcétera), o comentarios u opiniones implícitas expresadas en medios diversos, como redes sociales, en los que se puede tener en cuenta, a su vez, los tipos de relación entre los distintos usuarios y su fortaleza (*tie-strength* [2]). Por lo tanto, el principal reto de futuro es perfilar a cada usuario integrando información explícita e implícita, siendo la información explícita aquella que el usuario suministra directamente al sistema, y la información implícita aquella que el sistema infiere del comportamiento habitual del usuario, como puede ser el consumo de contenidos, la actividad o comentarios en una red social, etc.

2. Adquisición y modelado de la información del contenido

Por otro lado, es importante avanzar en la anotación y descripción de los contenidos. Dentro del tema de esta tesis, los contenidos abordados son imágenes y vídeos, lo que plantea retos añadidos para su caracterización. El problema habitual en cuanto a la descripción de los contenidos, especialmente en el caso de anotación y descripción automática, como el que se aborda en esta tesis, es la dependencia semántica del dominio de conocimiento. Por esa razón, existen dos vías fundamentales de investigación de este ámbito alrededor de los sistemas de recomendación. En primer lugar, presenta un gran auge el análisis semántico de los conceptos involucrados en el contenido, a través de la utilización de ontologías y otras tecnologías semánticas. Esta línea provoca que los recomendadores deban ser específicos para los dominios de conocimiento abordados [3]. Por otro lado, otra línea busca encontrar parámetros estructurales, morfológicos o de otro tipo que sean transversales al dominio de conocimiento y presenten rasgos comunes a un determinado tipo de media. En esta dirección, hasta ahora virgen en el terreno visual, sí existen investigaciones aplicadas al audio y, específicamente, a la música, como las propuestas de Zhi-Sheng [4], Reed [5] o Shao [6].

3. Mejora de la eficiencia y precisión de algoritmos de recomendación

Esta es la línea más específica de los sistemas de recomendación, ya que las anteriores presentan un cierto carácter transversal. Una posible clasificación de los sistemas de recomendación los agrupa según los tipos de técnicas utilizados [7]: técnicas basadas en contenido, técnicas sociales (también conocidas como de filtrado colaborativo, debido a la popularidad de esta técnica específica) y técnicas híbridas, que combinan las dos anteriores. En líneas generales, los nuevos sistemas tienden al empleo de técnicas de recomendación híbrida, lo cual incluye la necesidad de avanzar tanto en los algoritmos de recomendación basados en contenido, como en las técnicas de filtrado colaborativo y en las metodologías de hibridación.

4. Presentación de resultados y generación de confianza en el usuario

Por último, se empieza a dar mayor importancia a ciertos aspectos, aparentemente colaterales, que se habían dejado de lado para potenciar las técnicas de predicción. Estos aspectos se pueden resumir como la presentación de resultados al usuario y generación de confianza. Esto implica que las recomendaciones resulten claras y transparentes, que se permita al usuario conocer las razones por las cuales se le recomienda un producto, independientemente de la técnica utilizada para esta recomendación [8]. Además, el usuario debe de tener la opción de modificar los parámetros de la recomendación, de tal forma que siempre sea éste quien decida sus características, como por ejemplo, generar una recomendación más arriesgada pero más novedosa, u otra más conservadora pero más segura. Es decir, el objetivo es intentar una humanización de la máquina, de tal forma que la interacción hombre-máquina resulte no solo más agradable al uso de una persona, sino también más eficaz a la hora de comunicar las decisiones tomadas.

1.3 Aportes de la tesis

Esta tesis presenta aportes especialmente relacionados con los tres primeros puntos resaltados entre los retos del estado del arte, siempre centrados en los contenidos audiovisuales. Sin embargo, en todo momento se han tenido en cuenta los parámetros de usabilidad y confianza del usuario del punto 4, de modo que se han diseñado los algoritmos de forma que siempre sea posible una presentación de resultados y de explicaciones de la recomendación totalmente transparentes y parametrizables por el usuario. Sin embargo, los aportes concretos de investigación se han centrado en los 3 primeros puntos:

1. Mejora de la adquisición y modelado de la información del usuario

Se ha desarrollado un modelo de procesamiento automático de la información de consumo audiovisual de un usuario, de tal manera que se obtiene el interés de un usuario por un determinado ítem, tema o categoría semántica, dependiendo de las necesidades del sistema de recomendación. La descripción de este módulo corresponde al capítulo 4 de la tesis.

2. Mejora de la adquisición y modelado de la información del contenido

Se ha definido una serie de descriptores estructurales y morfológicos que permiten caracterizar imágenes y vídeos a través de su estética, y se ha comprobado su influencia en las preferencias de los usuarios. Estos descriptores, al estar directamente relacionados con las preferencias estéticas del usuario, se pueden aplicar a imágenes y vídeos con independencia de su contenido semántico. Del mismo modo, se han establecido las estrategias de anotación automática que posibilitan la aplicación de estos descriptores a los sistemas de recomendación. El modelo, junto con sus técnicas correspondientes, se presenta en el capítulo 5 de la tesis.

3. Desarrollo de algoritmos de recomendación más precisos, eficientes y adaptados al escenario

A partir de la información del usuario y del contenido obtenida en los dos puntos anteriores, se ha creado un nuevo algoritmo de recomendación basado en contenido y se han aplicado conocidas técnicas de hibridación que han permitido combinarlo con las técnicas tradicionales de filtrado colaborativo para poder realizar pruebas completas sobre un sistema híbrido. La descripción de estas técnicas corresponde al capítulo 6 de la tesis.

Estos aportes al conocimiento se han difundido a la comunidad científica a través de diversas publicaciones en congresos y revistas internacionales, como queda reflejado a lo largo de la tesis y, de manera explícita, en el apartado “Avance del conocimiento y divulgación” del capítulo 7 de la tesis.

El trabajo de la tesis doctoral se ha enmarcado en varios proyectos de investigación financiados por organismos nacionales. Concretamente, estos proyectos han sido:

- 3SME (Plataforma para el análisis semántico y medición del impacto de contenidos y clima social en un entorno multicanal. TSI-020302-2008-69. 2008-2009), en el que se comenzaron las investigaciones relativas al consumo audiovisual y a la captura y modelado de información de usuario en receptores de TV.
- JOI (Juegos Olímpicos Interactivos. TSI-020110-2009-430. 2009-2011), en el que se aplicó el modelo de inferencia de interés a partir del consumo a un escenario específico de Juegos Olímpicos y contenidos deportivos para crear un sistema de recomendación en cliente.
- BUSCAMEDIA (Hacia una adaptación semántica de medios digitales multirred-multiterminal. Programa CENIT-E. CEN-20091026. 2009-2012), en el que se abordó la adaptación de técnicas de recomendación a distintas arquitecturas de personalización, se desarrollaron los modelos de análisis y extracción de características estéticas de imágenes y vídeos, y se crearon los algoritmos de recomendación basados en contenido.
- MIREIA (Distribución de contenidos mediante Motores Inteligentes de Recomendación Adaptados a Redes Sociales. IPT-2011-2015-430000. 2011-2013), en el que se completaron los algoritmos de recomendación mediante la

hibridación de las técnicas ya existentes y se añadió al modelo de extracción de características de vídeo un soporte de atributos de audio.

El desarrollo de las investigaciones dentro de estos proyectos ha permitido, además, el desarrollo de prototipos utilizados para realizar las pruebas y validar así las técnicas y algoritmos creados.

1.4 Organización de la tesis

La tesis se estructura de acuerdo con la metodología seguida en el curso de la investigación y se divide en nueve capítulos de la siguiente manera.

El capítulo 1 introduce los principales temas de la tesis y los sitúa en su contexto.

El capítulo 2 expone los objetivos concretos de la tesis y describe la metodología seguida para la consecución de cada uno de ellos.

El capítulo 3 establece la manera en que las técnicas y algoritmos de recomendación, dependiendo de su naturaleza y tipología, tienen cabida dentro de sistemas de recomendación reales, al mismo tiempo que establece la arquitectura general en la que se encuadran los algoritmos diseñados en los tres capítulos siguientes.

El capítulo 4 describe las nuevas técnicas que permiten obtener información del usuario relacionada con sus preferencias a partir de su consumo audiovisual.

El capítulo 5 plantea un modelado de descriptores que permite caracterizar imágenes y vídeos, extrayendo de ellos información útil para los algoritmos de recomendación.

El capítulo 6 describe los algoritmos propuestos para generar las recomendaciones a partir de la información del usuario (capítulo 4) y del contenido (capítulo 5). Los capítulos 4, 5 y 6, que conforman el núcleo de la tesis, se corresponden con los tres aportes fundamentales indicados en el apartado anterior: “Mejora de la adquisición y modelado de la información del usuario”, “Mejora de la adquisición y modelado de la información del contenido” y “Desarrollo de algoritmos de recomendación más precisos, eficientes y adaptados al escenario”. Estos tres capítulos presentan avances específicos sobre el estado del arte, por lo que incluyen secciones autónomas de descripción del estado del arte y de pruebas y validación.

El capítulo 7 muestra unas conclusiones finales, incluyendo el resumen del avance del conocimiento que ha generado la tesis doctoral, así como las principales líneas de futuro que se dejan abiertas.

Finalmente, el capítulo 8 presenta las referencias bibliográficas empleadas en la elaboración de la tesis.

La tesis se completa con un capítulo de acrónimos y cuatro anexos que complementan datos puntuales referidos en los capítulos centrales.

2. Objetivos y metodología

El objetivo principal de la tesis es modelar sistemas multimedia para mejorar los procesos de personalización y recomendación de contenidos audiovisuales, tomando como punto de partida tanto los contenidos como el consumo audiovisual de los usuarios. Para ello, en esta tesis se generan nuevas técnicas, algoritmos y modelos de información que pueden ser integrados en arquitecturas completas de sistemas de recomendación multimedia, según la investigación previa sobre arquitecturas plasmada en el capítulo 3 de la tesis.

Este objetivo general se materializa en tres aportes concretos que avanzan sobre el estado del arte, y que quedan referidos en los apartados 2.1, 2.2 y 2.3 de este capítulo. Además, en cada uno de ellos se comenta la metodología que se ha seguido para su consecución, así como las particularidades que ha habido que tener en cuenta para poder demostrar la validez y mejoras de estas nuevas técnicas.

Estos tres objetivos concretos han surgido de las necesidades de los sistemas de recomendación ya comentadas en el apartado de introducción; por un lado, es necesario tratar la información disponible de los usuarios, por otro lado, modelar adecuadamente la información de los contenidos y, por último, resulta fundamental crear un algoritmo de recomendación adecuado a la naturaleza de los contenidos audiovisuales objeto del estudio. La Fig. 1 enmarca esta situación en un escenario genérico de distribución de contenidos audiovisuales.

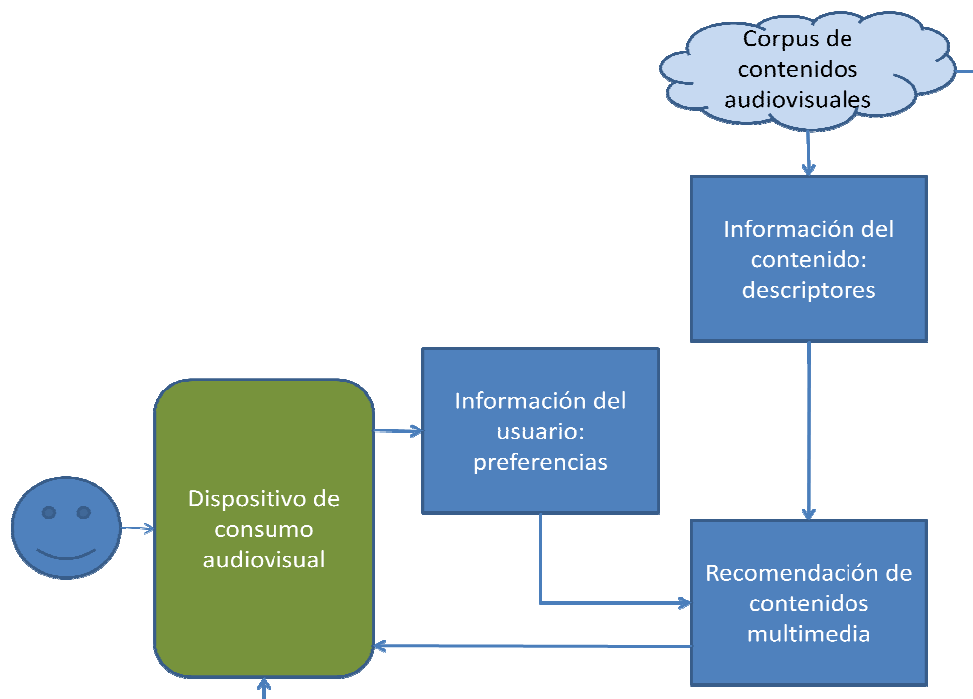


Fig. 1. Escenario general de distribución y recomendación de contenidos audiovisuales

2.1 Creación de un modelo que permita inferir preferencias de los usuarios a partir de su consumo audiovisual

2.1.1 Descripción

Los algoritmos de recomendación emplean, como una de las entradas, información explícita sobre las preferencias que los usuarios tienen acerca de determinados ítems o categorías. El objetivo que se propone en la tesis es obtener esa información de manera implícita, partiendo del consumo audiovisual del usuario, de tal forma que la interacción con el sistema de recomendación sea lo más transparente posible.

Como alternativa a los sistemas de recomendación tradicionales, en los que se solicita al usuario una valoración personal acerca del producto o contenido consumido, en esta tesis se plantea obtener esa información directamente a partir de la información de consumo del usuario de más bajo nivel. Para ello, se dibuja un marco en el que el usuario consume productos audiovisuales, de tal manera que se puede recabar directamente la información de interacción con el *player* del vídeo, que incluye el detalle de los instantes de visualización normal, parada, *fast-forward*, o repetición del consumo de determinadas partes del vídeo. Sobre esta información de bajo nivel, recogida a nivel de ítem, se genera una primera predicción sobre el interés que ese consumo denota. Posteriormente, sucesivas visualizaciones de contenidos relacionados permiten aumentar la confianza de las predicciones, de tal manera que un modelo matemático realimentado genera las predicciones finales acerca del gusto de un usuario por un ítem o por una categoría relacionada con ese ítem.

Estas predicciones deberían ser lo más cercanas posibles al gusto que el usuario declarara de manera explícita, de tal manera que un sistema basado en información implícita de consumo audiovisual pudiera ser equiparable a un sistema basado en información explícita. Esa información final de salida será la que emplee posteriormente el algoritmo de recomendación para la generación de predicciones de preferencias sobre ítems desconocidos por el usuario.

2.1.2 Metodología

Para desarrollar esta etapa de recomendación se establece la siguiente metodología:

- 1) Examinar la información de bajo nivel de consumo que se puede obtener del visualizador del vídeo, ya sea este un televisor o un *player* embebido en un navegador de Internet.
- 2) Examinar las posibilidades de relación de los contenidos audiovisuales, analizando tanto características comunes como categorización del tipo de contenidos.
- 3) Analizar el estado del arte relacionado con el modelado del consumo audiovisual dentro del marco de trabajo establecido.
- 4) Crear un modelo matemático que permita generar una primera predicción del gusto del usuario sobre el ítem que se ha consumido.

- 5) Desarrollar el modelo matemático completo que añada fiabilidad a la primera predicción y permita utilizar la información complementaria de consumos anteriores.
- 6) Validar el modelo mediante la realización de pruebas con usuarios reales. Para ello, será necesario:
 - a. Desarrollar o integrar un software de captura de datos de bajo nivel en un dispositivo de consumo audiovisual.
 - b. Elegir una serie de contenidos audiovisuales que estén disponibles para un conjunto acotado de personas
 - c. Elegir una muestra de usuarios y repartir entre ellos los dispositivos con el software de captura para obtener los datos de bajo nivel. Los usuarios utilizarán el dispositivo para consumir contenidos durante un tiempo previamente establecido.
 - d. Una vez cumplido el tiempo establecido, retirar los dispositivos a los usuarios y pasarles una encuesta en la que se pregunte, a cada uno de ellos, la valoración que daría, según su gusto personal, a determinados productos audiovisuales.
 - e. Aplicar el modelo matemático desarrollado a los datos de consumo recabados en el dispositivo para obtener la predicción de las valoraciones de los usuarios sobre los contenidos.
 - f. Comparar las valoraciones inferidas por el sistema con las valoraciones otorgadas de forma explícita por los usuarios a los contenidos mediante la encuesta. De esta manera se obtendrán las consecuentes medidas de precisión y bondad del ajuste de las predicciones.

El desarrollo de este objetivo y la explicación detallada de los pasos y pruebas necesarias queda recogido en el capítulo 4 de la tesis.

2.2 Creación de un modelo que permita describir imágenes y vídeos en función de factores influyentes en los gustos de los usuarios e independientemente de su contenido semántico

2.2.1 Descripción

Aparte de la información del usuario, la información de los contenidos es la otra fuente de conocimiento que necesitan como entrada los algoritmos de recomendación. Por esta razón, se plantea estudiar la naturaleza de los contenidos audiovisuales (imágenes y vídeos) y analizar las características de tipo formal y estético, relacionadas con parámetros morfológicos y estructurales que sirvan para determinar los gustos de los usuarios. Se plantea la definición de una serie de características independientes del dominio semántico del objeto audiovisual y que, por lo tanto, sirvan para cualquier dominio de conocimiento. Dado la mayor parte de las líneas de investigación actuales están centradas en el ámbito de la descripción semántica, se cubre aquí la investigación de una serie de parámetros de bajo nivel que también pueden ser determinantes a la hora de caracterizar las preferencias de los usuarios.

Por lo tanto, una vez que se haya establecido una hipótesis de parámetros de bajo nivel que puedan influir en las preferencias audiovisuales de los usuarios, es necesario

investigar estrategias de anotación automática, de tal manera que la descripción de los contenidos pueda ser automatizada dentro de un sistema completo de recomendación. Del mismo modo, es importante tener en cuenta la eficiencia de esta anotación, ya que, si bien en el caso de las imágenes no será restrictivo, sí será importante que la carga computacional no sea excesiva en el caso de los vídeos, dado que puede aportar beneficios importantes en las posteriores aplicaciones que esta anotación pueda realizarse en tiempo real. Finalmente, será necesario comprobar la hipótesis de descriptores y verificar, haciendo uso de bases de datos existentes (o, en su defecto, generando nuevas bases de datos), que los descriptores definidos determinan, efectivamente, las preferencias audiovisuales de los usuarios.

2.2.2 Metodología

En este caso, la metodología indicada se ejecuta dos veces, una primera para las imágenes y una segunda, completamente análoga, para los vídeos:

- 1) Analizar el estado del arte relacionado con la descripción y anotación automática de características y estudiar literatura relacionada con los procesos de percepción audiovisual del ser humano.
- 2) Realizar la hipótesis de descriptores de imágenes/vídeos influyentes en el gusto de los usuarios.
- 3) Buscar o generar una base de datos de preferencias de usuarios sobre imágenes/vídeos concretos predefinidos.
- 4) Investigar en el estado del arte técnicas y estrategias de anotación para crear un sistema de anotación automática de imágenes/vídeos.
- 5) Generar una base de datos de imágenes/vídeos que estén contenidos en la base de datos de preferencias de usuarios, y aplicar a estas imágenes/vídeos el sistema de anotación.
- 6) Cruzar la información de preferencias de los usuarios con los descriptores obtenidos de cada una de las imágenes/vídeos para verificar las hipótesis de influencia de los descriptores definidos. Para ello:
 - a. Estudiar las funciones de distribución de los distintos descriptores aplicados a las imágenes/vídeos y normalizar si procede.
 - b. Determinar el método estadístico de contraste de hipótesis, dependiendo de la naturaleza de las distribuciones
 - c. Establecer un procedimiento, cuantitativo o cualitativo, para relacionar la valoración numérica de los usuarios sobre los ítems con la distribución de los valores de los descriptores para la totalidad del corpus.
 - d. Aplicar el método elegido de contraste de hipótesis para determinar qué descriptores son influyentes y cuáles no.

El desarrollo de este objetivo y la explicación detallada de los pasos y pruebas necesarias queda recogido en el capítulo 5 de la tesis.

2.3 Creación de un sistema de recomendación personalizada, cuyos criterios se adapten dinámicamente a las preferencias de cada usuario.

2.3.1 Descripción

Una vez se dispone de la información de partida necesaria, tanto de los usuarios como de los contenidos, la última etapa del proceso de recomendación consiste en la aplicación de un algoritmo de predicción de valoraciones y decisión de ítems recomendados.

Con el objetivo de explotar el tipo de información obtenida en el objetivo 2.2, se plantea aquí una investigación alrededor de la recomendación basada en contenido que haga uso de los descriptores de tipo formal y estético definidos anteriormente. Para ello, se plantea la creación de un algoritmo capaz de recomendar a cada usuario según la particularidad de sus preferencias, es decir, teniendo en cuenta aquellos criterios o descriptores que se haya comprobado que afectan a su percepción y a su valoración de los contenidos audiovisuales. Por esa razón, el algoritmo debe personalizar el modelo general de descriptores previamente definido, particularizándolo automáticamente para cada uno de los usuarios y creando un subespacio vectorial de características personalizadas.

Como ocurre en los sistemas de recomendación, el uso de una sola técnica empobrece el resultado final [7], entregando resultados sesgados y poco configurables, por lo que la investigación en los algoritmos de recomendación basada en contenido se complementa con el uso de técnicas sociales. Esto implica la inclusión de un método de hibridación del citado recomendador basado en contenido con un algoritmo de filtrado colaborativo habitualmente utilizado en el estado del arte y en sistemas comerciales.

También es necesario resaltar que, tanto el nuevo algoritmo basado en contenido como la etapa de hibridación, deben respetar ciertas condiciones de usabilidad y confianza de los sistemas de recomendación. Por este motivo, los algoritmos generados deben permitir:

- 1) Ser totalmente configurables y parametrizables por el usuario, si bien esta parametrización debe considerarse opcional, y nunca obstaculizar una utilización básica del recomendador. Es decir, el algoritmo de recomendación debe permitir tanto tener un funcionamiento predefinido para cada usuario, como el ajuste de opciones avanzadas de configuración.
- 2) Ofrecer un grado de utilidad de las recomendaciones y una explicación de las razones que han llevado al sistema a generar la recomendación de un determinado ítem. Es decir, se debe garantizar una transparencia total de cara al usuario, sin perder de vista la usabilidad, que, en este caso, implica realizar una abstracción a alto nivel, fácilmente comprensible por cualquier usuario medio, de las razones matemáticas que hay detrás de una determinada recomendación.

Finalmente, resulta especialmente importante la validación de este objetivo, ya que incluye, implícitamente, la validación del conjunto de algoritmos de la tesis incluidos en el sistema de recomendación.

2.3.2 Metodología

Para desarrollar esta etapa de recomendación se establece la siguiente metodología:

- 1) Analizar el estado del arte de los algoritmos de recomendación basados en contenido
- 2) Crear un algoritmo basado en contenido a partir de los descriptores obtenidos previamente de las imágenes y los vídeos, que sea capaz de adaptarse, en cada ocasión, a las preferencias analizadas de cada usuario, modificando durante su ejecución los parámetros que considera. Para ello, es necesario diseñar dos etapas secuenciales:
 - a. Análisis automático de preferencias, que determine qué características influyen en el gusto de cada usuario y, en función de ellas, cree un subespacio vectorial personalizado.
 - b. Predicción de las valoraciones de los ítems desconocidos realizada sobre el subespacio personalizado del usuario
- 3) Buscar o crear un algoritmo de filtrado colaborativo análogo a los utilizados en el estado del arte o en sistemas comerciales.
- 4) Aplicar los algoritmos de recomendación a las bases de datos de preferencias de imágenes y vídeos ya creadas en la etapa 3) del objetivo 2.2.
- 5) Establecer las medidas de precisión y éxito de los algoritmos de recomendación y obtenerlas para los algoritmos y bases de datos propias.
- 6) Diseñar un algoritmo de hibridación que permita combinar los resultados del algoritmo de recomendación basado en contenido propio con los resultados del algoritmo de filtrado colaborativo de referencia.
- 7) Validar el algoritmo final utilizando las bases de datos fijadas como referencia, y comparar los resultados obtenidos por el sistema híbrido, con diferentes parametrizaciones, con los resultados individuales de cada uno de los algoritmos de recomendación.

3. Técnicas y arquitecturas de sistemas de recomendación multimedia

3.1 Modelado general y requisitos de los escenarios

Una de las principales restricciones de los sistemas de recomendación es la dependencia que tienen del escenario de aplicación, ya que los algoritmos pueden dejar de ser válidos, o resultar poco operativos, en determinadas circunstancias, dependiendo tanto de los datos de entrada disponibles, como de la manera de recabarlos o de los requisitos computacionales. A continuación se presentan los elementos restrictivos de cara al modelado de las arquitecturas, que se pueden agrupar como restricciones debidas a la naturaleza del contenido, restricciones relacionadas con la adquisición de información del usuario, y restricciones asociadas a las características del sistema y su caso de uso.

a) Restricciones debidas a la naturaleza de los datos del contenido

Una restricción fundamental a la hora de elegir las técnicas más adecuadas para el diseño de un sistema de recomendación viene impuesta por los metadatos disponibles o que se pueden extraer de los ítems que forman el corpus de recomendación.

El supuesto de una ausencia total de metadatos del contenido nos llevaría a la reducción a un sistema genérico de recomendación, en el que los ítems únicamente son representados mediante un identificador. De esta forma, solamente se podrían utilizar técnicas de recomendación sociales, que emplean la inteligencia colectiva de los usuarios de forma indirecta para generar las recomendaciones. Este es el caso más típico de sistemas de recomendación, en el que en los últimos años se ha producido un avance muy significativo en los algoritmos de filtrado colaborativo. Sin embargo, como ya se ha mencionado, ciertas deficiencias del filtrado colaborativo solo pueden subsanarse haciendo uso de técnicas alternativas, basadas en contenido. Además, existen escenarios en los que no se pueden utilizar técnicas sociales debido a que el canal de comunicación es unidireccional (caso de la difusión de televisión convencional) y no es posible combinar simultáneamente las preferencias de un gran número de usuario. Otro impedimento del filtrado colaborativo son los escenarios concurrentes o aquellos que no pueden soportar una alta carga computacional, como se expone en el apartado c.

Teniendo en cuenta, por estas razones, que es necesario buscar estrategias que mejoren la eficiencia de la recomendación basada en contenido, esta técnica también se ve completamente condicionada a la naturaleza de los datos disponibles de los contenidos. Tal y como se especifica en [10], se distinguen tres capas de metadatos asociados a un objeto multimedia, como se aprecia en la Fig. 2. La definición de estos metadatos se basa en el estándar MPEG-7 [11]

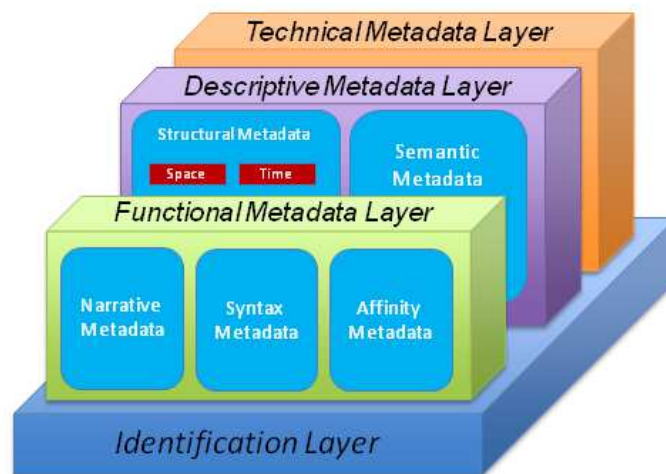


Fig. 2. Tipos de metadatos del objeto multimedia [10]

Los descriptores útiles para los algoritmos de recomendación son los recogidos en la capa de metadatos descriptivos y en la capa de metadatos funcionales. Dentro de la capa descriptiva, los metadatos semánticos dependen del dominio de conocimiento al que pertenezcan los objetos multimedia, por lo que no se consideran en esta tesis, por no poder establecer una estructura específica común. El resto de metadatos de la capa descriptiva (metadatos estructurales), así como los metadatos de la capa funcional, se tratan en profundidad en el capítulo 5 de esta tesis, estableciendo un modelo común para la recomendación que no depende del dominio.

La definición de la estructura común y escalable de metadatos llevada a cabo en el capítulo 5, permitirá el desarrollo de los algoritmos de recomendación basados en contenido que se explican en el capítulo 6, y que no son específicos de un sistema ni están asociados a un escenario determinado.

b) Restricciones relacionadas con la adquisición de información del usuario

Es importante tener en cuenta que la mayor parte de los sistemas de recomendación tradicionales cuentan con módulos de captura de información explícita, por lo que la mayoría de los estudios de sistemas de recomendación parten directamente de la información aportada por los usuarios acerca de sus gustos e intereses. De esta manera, se centran los esfuerzos de investigación en la manera de obtener recomendaciones más eficientes considerando esos datos de partida, que, además, al ser modelados de forma explícita, pueden reflejar directamente las características del usuario que el sistema deba conocer, ya sean cualitativas o cuantitativas.

Los problemas de este tipo de sistema son varios. Por un lado, los usuarios no siempre dicen la verdad, ya que pueden verse condicionados por factores familiares o sociales externos o, simplemente pueden no saber exactamente cuál es su preferencia, o el grado de interés que les provoca un determinado contenido. Además, la continua actualización de las preferencias por parte del usuario puede volverse tediosa y provocar su hartazgo, lo cual puede desvirtuar la validez de estas preferencias enunciadas explícitamente.

Por su parte, la captura de información implícita ha sido tradicionalmente descartada de cara a la recomendación, debido a la dificultad de realizar un modelado preciso, o se han realizado aproximaciones muy gruesas, como se especifica en el estado del arte del capítulo 4. Por lo tanto, el campo de investigación abierto en relación con el modelado de información implícita es realmente interesante en estos momentos, a pesar de los problemas que plantea.

Por estas razones, la tendencia actual es combinar las dos técnicas, utilizando tanto información explícita como implícita, al menos hasta que los estudios en el modelado de esta se encuentren más avanzados. En cualquier caso, lo importante es disponer de las herramientas que permitan la implantación de sistemas tanto de información explícita como implícita, de tal forma que esto no sea una restricción y el diseñador del sistema pueda elegir uno de los dos métodos o una combinación de ambos dependiendo de sus necesidades específicas.

En el caso que nos ocupa, de un sistema de recomendación de contenidos audiovisuales, se presenta una gran oportunidad para desarrollar investigaciones acerca del modelado de información implícita, muy propicio para el caso debido a que la idea es que el recomendador forme parte de una estructura ya definida de distribución de contenidos, con lo que se aspiraría a una máxima integración de los servicios.

El dispositivo de captura de datos, por estas razones, condiciona el tipo de información que se puede obtener, influyendo de esta manera en el modelado de los datos de cara a inferir el grado de interés y la posterior recomendación. Sin embargo, sea cual sea el dispositivo de consumo, ciertas características de bajo nivel siempre van a estar presentes, por lo que los algoritmos de inferencia del interés que se plantean en el capítulo 4 de la tesis son aplicables a cualquier tipo de dispositivo y de escenario, ya que, como se verá en ese mismo capítulo, su implementación podrá ser ejecutada tanto en un servidor como en el terminal del usuario.

c) Restricciones asociadas a las características y caso de uso del sistema

Por último, los requisitos no funcionales de un sistema también pueden determinar qué técnicas se pueden utilizar y cuáles no. En el caso de escenarios en los que es necesario repartir la carga computacional del recomendador en cada uno de los terminales cliente, quedan inhabilitadas las técnicas sociales, por lo que solo cabe la recomendación basada en contenido. En otros casos, en los que puede existir una cierta carga computacional en el servidor, pero con importantes limitaciones, es posible emplear técnicas sociales más ligeras que el filtrado colaborativo, como puede ser el uso de estereotipos, ya sean estos estáticos o dinámicos. Sin embargo, la recomendación más eficiente siempre se conseguirá haciendo uso de la mayor cantidad de recursos posibles, por lo que las técnicas definidas en esta tesis contemplan también la opción computacional menos restrictiva, en la que se pueden emplear sistemas híbridos para generar los resultados finales. Estas circunstancias provocan la aplicación de diferentes configuraciones a la arquitectura general, tal y como se expone en el siguiente apartado.

3.2 Adaptación a distintos escenarios

Debido a la dependencia de los factores del entorno, se presenta en este apartado la adaptación de los algoritmos que se desarrollan en la tesis a diversos escenarios de aplicación, cada uno de los cuales presenta restricciones específicas. Los tres escenarios planteados cubren las distintas peculiaridades de los tipos de sistema estudiados, y se resumen en la Tabla 1.

Arquitectura	Ejemplo de servicio	Naturaleza de los datos del contenido	Adquisición de las preferencias de usuario	Tipos de técnicas aplicables
Recomendación en cliente	Radiodifusión TV	Metadatos conceptuales asociados	Implícita a partir del consumo.	Recomendación basada en contenido
Recomendación en cliente o con baja carga en servidor	Distribución vídeo Internet concurrente	Metadatos técnicos, estructurales y funcionales	Implícita a partir de consumo. Explícita: valoraciones sobre los contenidos	Recomendación basada en contenido. Técnicas sociales de baja carga
Recomendación en servidor	Búsqueda de contenidos	Opción 1: ausencia de metadatos. Opción 2: metadatos técnicos, descriptivos y funcionales	Implícita a partir de consumo. Explícita: valoraciones sobre los contenidos	Híbridas: Recomendación basada en contenido y técnicas sociales

Tabla 1. Características de los escenarios de aplicación

3.2.1 Arquitectura con recomendación en cliente

Ejemplo: escenario de radiodifusión de televisión digital convencional

Un escenario de radiodifusión de televisión convencional puede presentar distintas características. La más importante de ellas a la hora de definir un sistema de recomendación es la presencia de un canal de retorno que permita comparar en el lado servidor información de consumo y preferencias de distintos usuarios. Este caso permitiría el desarrollo de un sistema de recomendación basado en técnicas sociales, ya que no tendríamos restricciones de ese tipo.

Sin embargo, dadas las características del mercado, sería posible, por ejemplo, que un fabricante de set-top-boxes deseara implementar un sistema de recomendación que estuviera íntegramente desarrollado en el terminal, sin depender de una conexión externa, ni de la presencia de otros usuarios. Para ello, dado que no podría emplear técnicas sociales, debería hacer uso de los metadatos que los operadores envían junto

con la señal audiovisual para poder desarrollar un sistema de recomendación basado en contenido. Además, dado que un set-top-box es un dispositivo creado fundamentalmente para consumir contenido audiovisual, sin llevar a cabo una gran interacción con el usuario, se impone también la restricción de que todas las preferencias del usuario sean obtenidas de forma implícita a partir del consumo audiovisual.

Por estas razones se plantea un modelo de recomendación en cliente a partir del consumo audiovisual del usuario y de los metadatos radiodifundidos por los operadores, siguiendo la arquitectura de la Fig. 3.

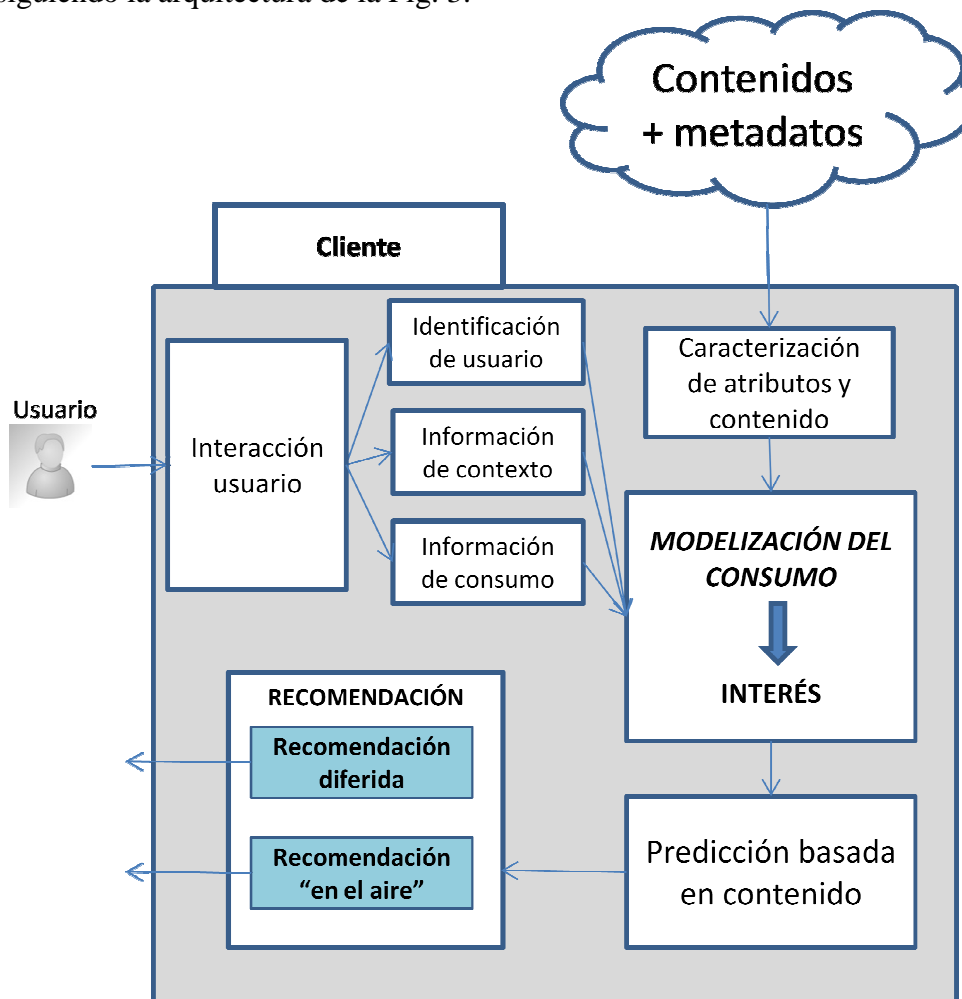


Fig. 3. Arquitectura en un escenario *broadcast*

Como dispositivo de usuario se ha trabajado en las pruebas con receptores de televisión digital terrestre, dotados de capacidades interactivas para poder realizar desarrollos sobre ellos. Sin embargo, la arquitectura propuesta se podría adaptar a cualquier otro dispositivo con las capacidades suficientes para recibir contenidos audiovisuales y desplegar aplicaciones de control o interacción, como podría ser un televisor digital, un *smartphone*, una *tablet-PC*, etcétera.

El hecho de que el módulo de modelización del consumo transforme las medidas de consumo audiovisual en una estimación cuantitativa del interés que un ítem o una categoría despierta sobre un usuario aporta la flexibilidad necesaria para que esta técnica de inferencia de información implícita de usuario se pueda aplicar tanto a técnicas de recomendación basadas en contenido como a técnicas de recomendación

social, como el filtrado colaborativo, que emplea *ratings* (valoraciones numéricas de los usuarios) como punto de partida para los algoritmos.

La arquitectura planteada en la Fig. 3 parte de la recogida de las interacciones de consumo del usuario, lo cual implica una previa identificación del usuario y una posible recogida implícita de otro tipo de datos contextuales (por ejemplo, si se han identificado simultáneamente varios usuarios, se sabe que están viendo la televisión acompañados, lo cual es un factor que se tiene en cuenta a la hora de inferir el interés, con su grado de fiabilidad asociada, como se verá en el capítulo 4 de la tesis). Esta información, junto con los metadatos radiodifundidos que identifican los contenidos y los relacionan mediante categorías o características determinadas, se emplea en la etapa de modelización del consumo para inferir las preferencias de los usuarios. Llegados a este punto, la técnica elegida para la recomendación condiciona que esa información de interés se envíe de vuelta al servidor para lanzar los algoritmos de filtrado colaborativo (u otras técnicas sociales) o permanezca en el dispositivo del usuario para alimentar el módulo de recomendación basado en contenido. Ambas opciones son planteadas en [9]. En este caso, el algoritmo de recomendación se simplifica, y únicamente decide, a partir de las similitudes entre las preferencias inferidas y los metadatos de los próximos contenidos que se van a emitir, qué se recomienda, en qué grado, y por qué razones.

Como ya se ha comentado, existen dos variantes para los sistemas de recomendación basados en contenido desarrollados en cliente: aquellos que emplean únicamente la relación temporal entre elementos aislados o, en todo caso, una relación suave mediante categorías o tipos de programa, como se explica en [9], y aquellos otros que definen cada uno de los contenidos mediante una serie de atributos, en los que el papel del módulo de inferencia de interés se extrapola del contenido a los atributos que lo definen, tal y como se explica en [12]. En el capítulo 4 de la tesis se entra en los detalles tanto del modelo creado para la inferencia del interés como de su aplicación directa a sistemas de recomendación basados en contenido.

3.2.2 Arquitectura con recomendación en cliente o baja carga en servidor

Ejemplo: escenario concurrente con recomendador en cliente

Un escenario similar al anterior se plantea dentro de una red como Internet en la que se realice distribución de vídeo y en la que haya una alta concurrencia de usuarios, provocando esto que la carga computacional en el servidor deba ser reducida. De esta forma, no se excluyen las técnicas de recomendación social, pero es necesario pensar alternativas al filtrado colaborativo, debido a su alto coste de recursos. Una posible arquitectura se muestra en la Fig. 4

Dado que el escenario está más adaptado a una mayor interacción del usuario, se contempla la posibilidad de recoger también información explícita de las preferencias, tales como valoraciones de los usuarios sobre los contenidos consumidos. De esta forma, se podrían utilizar valoraciones explícitas cuando estas existan, pero se emplearía el sistema de inferencia de interés a partir del consumo para caracterizar el gusto de los usuarios sobre el resto de contenidos consumidos, de tal forma que no sea necesaria la participación activa del usuario para el funcionamiento del recomendador, y

esto solo sea una opción más. Como se ha comentado en capítulos precedentes, es importante no forzar al usuario a realizar determinadas interacciones y, desde luego, no supeditar el funcionamiento del sistema a estas interacciones. Por esta razón, se concluye que el modelo híbrido es el más interesante, ya que además permite ajustar mediante la información explícita ciertos parámetros del módulo de inferencia de interés, y comprobar la eficiencia de su funcionamiento al comparar los resultados de ambas vías de extracción de interés.

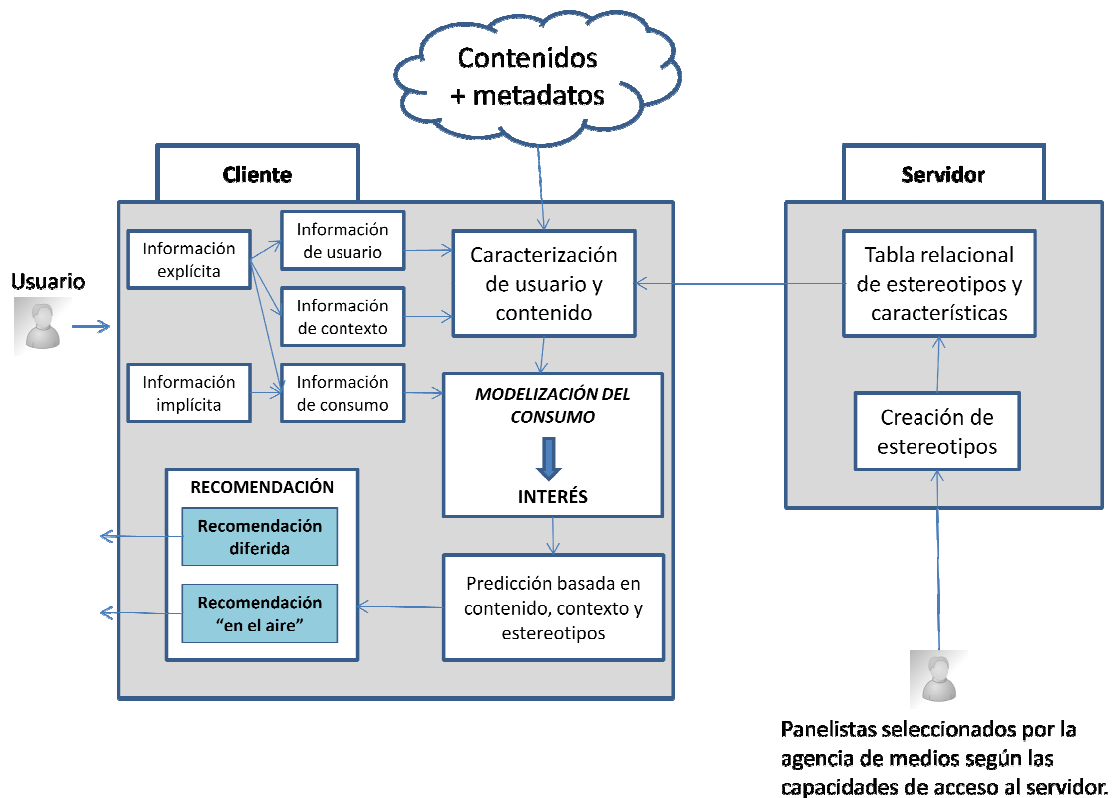


Fig. 4. Arquitectura en un escenario concurrente

Debido a la peculiaridad señalada sobre los requisitos computacionales, la mayor inteligencia del recomendador está en la parte cliente, que ya no se limita solamente a una captura de la información. Sin embargo, resulta interesante no prescindir totalmente del servidor, con el fin de no perder la información derivada de la inteligencia aportada por el "colectivo social".

De este modo, se plantea como solución utilizar un *clustering* de usuarios mediante estereotipos inferidos de una muestra de población estadísticamente elegida, a la manera de un panel de audimetría. Sobre esta muestra de usuarios, se relacionarían sus consumos con las características de los contenidos, con lo que se obtendría una relación entre estereotipos y sus características.

Así, en el cliente, además de toda la parte de recogida y modelado de datos de consumo, se puede realizar una caracterización de los usuarios y sus preferencias utilizando estos estereotipos. También quedaría en el cliente el motor de recomendación en sí, que se basaría fundamentalmente en contenido, pero que emplearía la parte social descrita a través de la metodología de los estereotipos, con el fin de obtener una recomendación final más fiable y certera. Además, el recomendador basado en contenido podría ser más desarrollado y ampliarse a cualquier tipo de contenido, ya que el uso de Internet

facilitaría el acceso a metadatos más complejos o extraídos previamente del contenido, según la caracterización que se desarrolla en el capítulo 5 de esta tesis.

3.2.3 Arquitectura con recomendación en servidor

Ejemplo: escenario orientado a búsqueda

Por último, se presenta otra posible aplicación a un escenario Web (Fig. 5), pero en este caso no directamente de consumo, sino de búsqueda de contenidos (con vistas, eso sí, a un inmediato consumo).

En esta arquitectura se presenta un número indefinido de usuarios que interactúan con una plataforma de búsqueda dotada de funcionalidades completas de distribución y visualización de contenidos multimedia. Los usuarios interactuarán con la plataforma a través de un PC o de un dispositivo personal. Este dispositivo llevará un software capaz de obtener información del usuario y del contexto y medir con suficiente precisión los consumos que se llevan a cabo. Finalmente, se realizará una comunicación con el servidor para transferir esos datos.

El servidor recogerá los datos de todos los usuarios. Recibirá de cada dispositivo personal la información de consumo y la introducirá en una base de datos. A continuación, el conversor de consumo a interés actualizará esta base de datos de usuarios con el valor de interés inferido.

Por otro lado, cuando se desee obtener una recomendación, los diversos módulos establecidos podrán utilizar directamente la información de la base de datos para ejecutar sus algoritmos de predicción. Sobre estos datos, se recomendarán al usuario los elementos con una mayor predicción de interés.

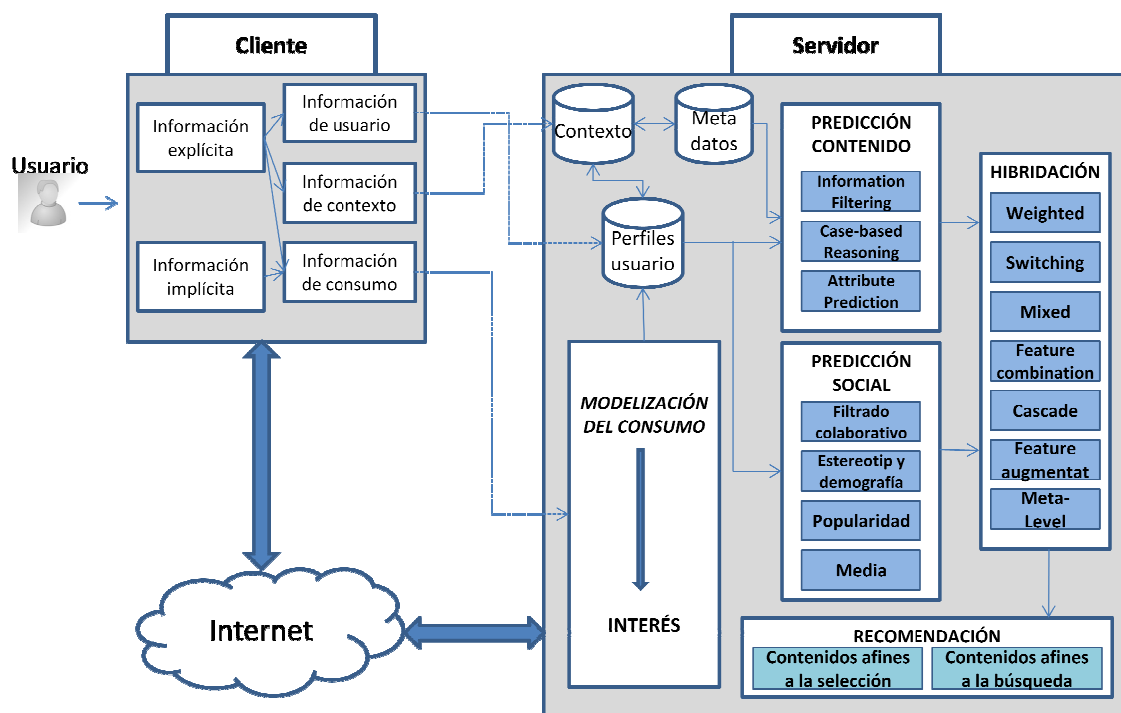


Fig. 5. Arquitectura orientada a búsqueda

En este caso se supone que no está limitada la carga del servidor o servidores, por lo que se pueden implementar todas las técnicas de recomendación posibles, tanto sociales como basadas en contenido, sin imponer restricciones no funcionales. La Fig. 5 recoge los tipos de técnicas más populares recogidas en el estado del arte dentro de la clasificación general de técnicas basadas en contenido (*Information Filtering*, *Case-Based Reasoning*, *Attribute Prediction*), técnicas sociales (*Filtrado colaborativo*, *Estereotipos y demografía*, *Popularidad*, *Media*) y técnicas de hibridación (*Weighted Switching*, *Mixed*, *Feature combination*, *Cascade*, *Feature Augmentation*, *Meta-Level*) [13].

La heterogeneidad de los tipos de búsqueda y las posibilidades que da una capacidad computacional distribuida y con bases de datos que puedan utilizar gran cantidad de información, permite que se desarrollen módulos de recomendación de complejidad creciente, obteniendo en las diferentes combinaciones posibles de hibridación una solución que puede ser muy eficiente. Del mismo modo, el resultado de la recomendación puede orientarse hacia el usuario con un carácter general o particular.

En la primera opción, en función de la búsqueda y del historial del usuario obtenido con anterioridad, el sistema realiza las recomendaciones que más se ajusten a lo que el usuario pide. En la segunda, de carácter particular, se enfoca la recomendación hacia la similitud entre contenidos partiendo de lo que el usuario ha consumido, es decir, a partir de la selección de contenidos que ha realizado sobre los resultados recomendados. Este método podría ser iterativo e ir redimensionándose en tiempo real.

La amplia infraestructura involucrada en un motor de búsqueda, con una gran cantidad de metadatos asociada a cada uno de los objetos multimedia indexados, permite la utilización de algoritmos de recomendación muy variados.

Siguiendo la definición de metadatos mostrada en la Fig. 2, se explica en [10] una estructura de metadatos diferenciada en tres capas:

- 1) La capa de metadatos técnicos es útil para funcionalidades de búsqueda y para realizar filtrados específicos o aportar información complementaria al usuario, pero en esta tesis no se consideran para generar directamente recomendaciones a partir de ellos, ya que no se trata de atributos intrínsecos al objeto audiovisual, sino de características de una determinada réplica. Ejemplos de metadatos técnicos puede ser el formato de un vídeo, su códec de vídeo o audio, la resolución, tasa binaria, etc.
- 2) La capa de metadatos descriptivos engloba tanto metadatos estructurales como semánticos. Ambos tipos pueden resultar muy útiles para la recomendación. Sin embargo, en esta tesis sólo se hará uso de metadatos estructurales, dado que no son específicos de determinados dominios de conocimiento y son comunes a un determinado tipo de media. Ejemplos de metadatos estructurales pueden ser los que hacen referencia a la luminosidad o a la textura de una imagen, mientras que los metadatos semánticos articulan conceptos relacionados con el contenido de esa imagen (mediante el reconocimiento de objetos y el posterior uso de ontologías para relacionar conceptos, por ejemplo).
- 3) La capa de metadatos funcionales es aquella construida a partir de los metadatos que los distintos procesos funcionales (búsqueda, recomendación, generación automática, etc.) crean dentro de sus procesos y a partir de los metadatos de

capas inferiores para llevar a cabo sus operaciones. El almacenamiento de estos metadatos permite enriquecer la información relativa a un objeto multimedia, con la consiguiente posibilidad de mejorar el rendimiento e incorporar facilidades a otros procesos no directamente relacionados. Algunos ejemplos pueden ser los metadatos sintácticos (como podría ser la duración media de plano de un vídeo), los metadatos narrativos (como la densidad de escenas), o los metadatos de afinidad (como la media de valoraciones de los usuarios sobre el objeto multimedia). La definición de metadatos útiles para la recomendación se desarrolla ampliamente en el capítulo 5 de la tesis.

La heterogeneidad de los procesos involucrados en la indexación y búsqueda de contenidos multimedia abre una gran cantidad de posibilidades a la recomendación multimedia, cuyo tratamiento dentro de esta tesis han sido acotados a las funcionalidades ya descritas. Por lo tanto, en este escenario se pueden emplear todos los mecanismos de personalización presentes en el estado del arte, y todos aquellos que vayan más allá en la comprensión de la percepción y las preferencias del ser humano.

Los tres escenarios expuestos presentan posibles aplicaciones, con diferente nivel restrictivo, a los avances en el estado del arte que se presentan en esta tesis. Por lo tanto, los siguientes capítulos de la tesis explican, en primer lugar, la manera de obtener preferencias a partir del consumo audiovisual de los usuarios (capítulo 4), a continuación, la extracción de descriptores relevantes de los contenidos (capítulo 5) y, por último, el desarrollo de algoritmos que generen recomendaciones a partir de esos *ratings* y descriptores (capítulo 6).

A modo de conclusión, la Tabla 2 sintetiza la posibilidad de aplicar algunas de las técnicas descritas a los tres tipos de arquitectura propuestos.

Técnica	Recomendador en cliente	Recomendador en cliente o con baja carga en servidor	Recomendador en servidor
Inferencia del interés del usuario a partir del consumo audiovisual	✓	✓	✓
Recomendación basada en contenido	✓	✓	✓
Recomendación basada en estereotipos	✗	✓	✓
Recomendación basada en filtrado colaborativo	✗	✗	✓

Tabla 2. Aplicabilidad de técnicas a los distintos tipos de arquitectura

Es necesario recordar que, dado que la tesis se centra en recomendación de contenido audiovisual y las técnicas desarrolladas analizan sus características, los escenarios que se han planteados son escenarios controlados, en los que se tiene constancia de qué elementos existen en el corpus para ser recomendados, debido a un procesamiento previo que puede ser manual o automático. Por esta razón, la recomendación basada en

contenido se puede aplicar en todos los casos; por el contrario, ante un escenario completamente abierto en el que no existiera ni se pudiera extraer ninguna información de los contenidos, ya que estos podrían ser de cualquier naturaleza (no necesariamente audiovisual) no se podría recurrir a las técnicas basadas en contenido.

La naturaleza de un contenido, de cara a la recomendación, viene dada por su descripción disponible antes que por su esencia; es decir, si la única información disponible de un vídeo es un identificador, ligado con valoraciones de distintos usuarios sobre él, carece de importancia de cara a la recomendación el hecho de que lo que queremos recomendar sea un vídeo y tenga, por lo tanto, naturaleza audiovisual. Si se tratara, por ejemplo, de un libro, ese mismo tipo de información podría estar disponible y la técnica de recomendación (siempre ligada al tipo de información disponible) podría ser la misma. Esta es la razón por la que, en una tesis focalizada en la recomendación de contenidos audiovisuales, es necesario considerar escenarios controlados, en los que la ingesta de contenidos esté asociada a procesos de anotación predefinidos.

4. Modelado del consumo audiovisual del usuario para la obtención del interés

4.1 Introducción

Una vez estudiados los posibles escenarios de aplicación y distintas arquitecturas para el desarrollo de sistemas de recomendación de contenidos, se llega a la conclusión de que todas ellas, ya sea de manera directa o indirecta, hacen uso de algún tipo de medida del gusto o interés de cada usuario por un determinado tema o un contenido. Estos serían los datos de entrada de usuario necesarios para el sistema, que quedan almacenados en los perfiles de usuario o relacionados con los mismos a través de alguna base de datos.

La obtención de estos datos referentes al interés puede realizarse de manera implícita o explícita, siendo este último método el más habitual entre los actuales sistemas de recomendación, en los que se pregunta al usuario una valoración concreta de aquello que ha consumido. Sin embargo, la captura y modelado implícito de los datos, todavía incipiente en el desarrollo del estado del arte del tema, como se verá en el apartado 4.2, reviste un interés especial por varias razones:

- 1) **Transparencia para el usuario.** Importante característica, dado que la información implícita permite desarrollar recomendadores en los que no sea necesaria la participación activa del usuario, pudiendo adaptarse así a diversos tipos de sistemas que no permiten tal participación y, en todo caso, pudiendo añadirse a una captura de información explícita, complementándola. Utilizando información implícita, además, se evita imponer esta participación activa del usuario como requisito imprescindible para el funcionamiento del sistema.
- 2) **Objetividad.** Una captura de información explícita es más fácilmente manipulable por un usuario, que puede “engañar” al sistema en un sentido o en otro, a causa de diversas razones socioculturales, o dependientes de la naturaleza concreta del sistema de recomendación. Al realizar un modelado invisible para el usuario, este sólo podría manipular la información auténtica modificando sus hábitos de conducta y, por lo tanto, provocando un perjuicio para él mismo. Esta característica quizás no tenga relevancia para el propio usuario, pero puede tenerla para los distribuidores de contenidos o para otros actores de la cadena de valor.

Al ser menos directa y más difícil de tratar, el estudio de la información implícita ha sido menos tratado en la literatura que la obtención de información explícita, a pesar de que el campo de investigación de esta última sea mucho más reducido. El objetivo final siempre resulta ser la obtención del valor de interés del usuario más cercano a la realidad. El escaso desarrollo de metodologías de modelado implícito que trabajen fuera

del ámbito semántico, así como mayor la complejidad en la posterior integración en los sistemas, ha llevado a los sistemas comerciales a recurrir a la información explícita, ya sea a través de formularios o preguntas concretas realizadas “en directo”.

La información implícita proviene de todos aquellos comportamientos observables que puedan ser recogidos por un sistema, sin intrusión hacia el espectador, y aprovechando sólo aquellas variables de las que éste haga uso. Por este motivo, **se hace necesario un modelado matemático de la información, que perseguirá inferir el interés del usuario en un determinado contenido o tema** que debe ser lo más aproximado posible a la realidad, por encima incluso de la certeza que el usuario crea tener de su propio interés. Sin embargo, la comprobación de los modelos estadísticos, dentro de la aplicación a sistemas de recomendación, debe ir encaminada al ajuste de la información de interés inferida con la información aportada por el usuario considerando este conocimiento del usuario como el límite de éxito objetivo. Este debe ser el límite marcado en un trabajo científico de estas características, ya que, una vez conseguido este objetivo, la pregunta para el futuro sería averiguar si puede el sistema saber más que el propio usuario acerca de sí mismo. Esta pregunta queda fuera del alcance de este trabajo, ya que linda con otros campos del conocimiento y está más próxima a estudios de carácter filosófico o psicológico, pero queda abierta para una futura línea de trabajo interdisciplinar.

Organización del capítulo

Este capítulo, que se centra dentro del esquema general de la tesis en la adquisición de información de usuario, parte de un estudio del estado del arte que es independiente del de los sistemas de recomendación, si bien se puede aplicar a estos, recogido en el punto 4.2.

A continuación, en el punto 4.3, se formalizan los datos de entrada del usuario necesarios para realizar el modelado del consumo. Estos datos se obtienen de la interacción del usuario con su dispositivo personal de consumo. A partir de estos datos formalizados, en el punto 4.4 se describe el modelo completo, que se desgana en dos etapas (obtención del interés local e inferencia final del interés global).

Los dos siguientes apartados, 4.5 y 4.6, aplican el modelo teórico general a dos escenarios diferentes: entornos categorizados y entornos multiatributo, respectivamente. De esta manera se demuestra la versatilidad del modelo y su posible adaptación a entornos diversos. Una vez descritos los procesos de adaptación a ambos escenarios, en el apartado 4.7 se explican en profundidad las pruebas realizadas para la validación, tanto para entornos categorizados (4.7.1) como multiatributo (4.7.2).

Finalmente, a modo de conclusión, el apartado 4.8 resume los resultados obtenidos.

4.2 Estado del arte

El modelado del consumo audiovisual destinado a sistemas de recomendación ha sido un tema tradicionalmente muy poco tratado en el estado del arte, debido, sobre todo, a que la mayoría de estos sistemas utilizan datos obtenidos de manera explícita, los cuales son totalmente adaptables a los requisitos necesarios, a modo de cuestionario o encuesta, sin necesidad de ninguna clase de modelado o procesamiento automático.

Sin embargo, se han encontrado algunas referencias más o menos directas en diversos artículos científicos que persiguen obtener una medida de interés (*rating*) de manera automática a partir del conocimiento de las acciones del usuario sobre un contenido audiovisual.

Las aproximaciones más generales y flexibles, que se pueden adaptar a diversos escenarios de explotación, están reflejadas en los artículos de Hyoseop et al. [14] y Zhiwen et al.[15]. Ambos artículos presentan estudios de sistemas de recomendación y personalización de contenidos en televisión, empleando métodos de inferencia implícita del grado de interés.

Hyoseop define como dato de entrada, además del consumo en sí mismo (al cual define como el principal y el que serviría para cualquier dispositivo), otras acciones como el almacenamiento o el marcado. Estos dos últimos tipos de acciones son aplicables a dispositivos de grabación, como un PVR (*Personal Video Recorder*), en los que es posible tanto almacenar un contenido como dejarlo marcado en forma de favorito.

De esta manera, Hyoseop establece el grado de interés sobre un contenido d mediante la fórmula expresada en (1).

$$R_d = \frac{\sum_{i=1}^m c_i}{T_d} + T_d \cdot \sum_{j=1}^n s_j + \sum_{k=1}^p a_k \quad (1)$$

Donde c_i son los tiempos parciales de consumo para cada una de las m acciones de consumo observadas durante el tiempo de duración del ítem, T_d . Por esta razón, un consumo continuo del tiempo de duración del propio programa equivaldría a una sola acción de valor 1. Este consumo se complementa con las n acciones de almacenamiento, que llevan asignado un peso s_j , ponderado según el tiempo de duración de esa grabación. Por último, a_k asigna pesos a las p acciones de marcado o agregación del contenido a favoritos.

La manera de emplear este modelo en diversos escenarios es combinar la influencia de los tres términos, empleando en cada caso aquellos cuya información se conozca, o que puedan ser ejecutados por un usuario. Sin embargo, el único término que es común a todos los escenarios de consumo audiovisual es el primero, por lo que en la mayoría de los casos la fórmula se reduce a la expresión de (2), en la que se advierte que la aproximación es únicamente una traslación lineal de tiempo de consumo en interés.

$$R_d = \frac{\sum_{i=1}^m c_i}{T_d} \quad (2)$$

Este modelo, aunque puede ser adaptado a diversos escenarios, resulta más útil en entornos en los que se dispone de realimentación directa y específica sobre el consumo de un contenido audiovisual y, en cualquier caso, se abre un importante campo de investigación para aprovechar la ingente cantidad de datos que los nuevos sistemas de visualización permiten recabar.

Por otro lado, existen otras aproximaciones al problema más sofisticadas, pero enfocadas a escenarios más concretos, focalizadas en dominios de conocimiento específicos, o destinadas a cubrir determinadas interacciones de los usuarios. Sin

embargo, estas ideas sirven de partida para realizar una reformulación de los casos y encaminar la investigación a partir de la base de Hyoseop y Zhiwen.

Liang et al. [16] propone un sistema para contenidos deportivos basado en eventos y protagonistas, y en la anotación automática de los mismos. De este modo, el interés de un usuario viene determinado por la presencia de uno de sus jugadores favoritos, o por la ocurrencia de un evento hacia el que el usuario muestre predisposición (por ejemplo, un gol en un deporte como el fútbol).

Más generales respecto al dominio de conocimiento son los estudios de Syeda-Mahmood y Ponceleon [17], que contemplan el aprendizaje de los gustos de los usuarios para realizar vistas previas de vídeos en un contexto de búsqueda. Para ello, realizan el modelado mediante un proceso oculto de Markov (HMM), en el que los estados ocultos representan el tipo de interés que determinado vídeo automáticamente generado provoca en el usuario (Tabla 3). Para medir este interés, el HMM utiliza como parámetros observables una serie de eventos, como la reproducción, pausa, deslizamiento *frame a frame*, avance rápido, avance lento, retroceso rápido y retroceso lento. Por lo tanto, aquí se alcanza una aproximación cualitativa al interés de los usuarios, aunque enfocado a un escenario concreto (búsqueda, que se podría aplicar, bien a Web, bien a dispositivos de almacenamiento como PVRs), que permita realizar acciones concretas sobre un vídeo que se visualice de forma diferida.

State	Description
S1	Courious
S2	Aimless browse
S3	Found something interesting
S4	Abandoned interest
S5	Looking for something
S6	Resumed interest
S7	Found what I wanted
S8	Undetermined
S9	None of the above
S10	Did not find what I wanted

Tabla 3. Descripción de los estados [17]

En el caso de la recomendación de contenidos, el interés inferido debe ser mostrado de forma cuantitativa, no cualitativa (a no ser que se trate de una recomendación puramente contextual), de tal modo que exista una escala ordenada sobre la que se puedan desarrollar los algoritmos de recomendación. Sin embargo, resulta muy interesante combinar valoraciones cuantitativas y cualitativas, especialmente en aquellos sistemas que puedan disponer de información de contexto.

Por último, Zimmerman et al. [18] se sitúa en un escenario más próximo al de la recomendación en directo, en el que ya no es necesario trabajar directamente sobre el contenido, sino únicamente tener conocimiento sobre si un determinado evento ha sido o no consumido. El modelo resultante es bastante rígido, ya que no infiere el interés, sino que trabaja directamente con recomendación sobre consumos: decide si un usuario, según sus comportamientos pasados, consumirá o no un determinado contenido, imponiendo una decisión rígida y binaria. Para ello, emplea un clasificador bayesiano, definido en 1998 por Pazani [19] y anticipado en 1973 por Duda y Hart [20], y árboles

de decisión, basados en los estudios previos de Quinlan en 1983 y 1991 [21]. Para el diseño del sistema, se deciden reglas de actuación a partir de la metadatos de los contenidos y la información almacenada en los perfiles de usuario (ejemplo en Tabla 4). Sin embargo, la flexibilización de este modelo puede ser uno de los grandes retos para la investigación en el campo del modelado audiovisual, ya que actualmente crecen las aplicaciones horizontales, que no distribuyen el contenido, con lo que la información de consumo con la que se cuenta es limitada y procede de fuentes externas, y no de la clásica realimentación directa a partir de la visualización de un contenido audiovisual.

Feature	Watched (C+)	Not-watched (C-)
Total programs	66	66
Station: ABC	10	56
Station: BBC	20	10
Title: BBC News	10	2
Actor: Jim Carrie	2	0
Genre: comedy	1	7
Keyword: murder	5	1

Tabla 4. Ejemplo del sistema de Zimmerman [18]

4.3 Captura de datos de entrada de fuentes heterogéneas (información explícita e implícita)

Los datos de entrada del sistema referidos a las medidas del consumo audiovisual del usuario deben permitir caracterizar completamente el proceso de consumo audiovisual de un usuario.

La variedad de dispositivos en los que se desarrolla actualmente el consumo audiovisual complica la tarea de caracterización, por lo que es necesario definir parámetros que sean independientes del tipo de terminal del usuario.

En primer lugar, es necesario especificar el tipo de *media* para el que se establece el modelo de captura de datos. Aunque la tesis se refiere a la recomendación tanto de imágenes como de vídeos, para el caso de consumo audiovisual se contempla solo el estudio de los vídeos, por su especial interés. El consumo de una imagen es instantáneo y se podría traducir simplemente en una métrica binaria, por lo que no se considera en este estudio. Aunque sería posible establecer métricas de impacto según el tiempo de retención de atención que opera una imagen sobre un usuario, este análisis presenta complejidades de tipo psicológico y perceptivo que quedan fuera del estudio de esta tesis. Por lo tanto, a continuación se va a proceder a la parametrización de los datos que caracterizan el consumo de vídeo por parte de un usuario.

La captura de datos implícitos con el fin de realizar una recomendación es un tema apenas tratado directamente en el estado del arte, pero es posible emplear técnicas y metodologías que pudieran tener unos propósitos iniciales distintos. Concretamente, y dado que en este trabajo se plantea obtener recomendaciones a partir de los consumos de los usuarios, se ha detectado una importante analogía con los sistemas clásicos de medición de audiencias. Esto resulta lógico, ya que tanto un sistema de recomendación como otro de medición de audiencias utilizan la información de consumo audiovisual de los usuarios, cuantificados y estructurados dentro del dispositivo cliente de interacción.

La naturaleza de este dispositivo cliente determina la manera en que se diseña e implementa la captura de datos pero, en un escenario como el actual, que tiende a la convergencia multirred y multiterminal, parece que lo indicado es diseñar sistemas que permitan una sencilla adaptación a múltiples dispositivos. En este sentido, resultan muy interesantes algunas investigaciones del campo de la medición de audiencias. Resulta representativo, en este sentido, el estudio de Álvarez et al. [22], que presenta un escenario de medición de audiencias convergente a diferentes redes y terminales.

Dentro de este escenario, en primer lugar, es necesario especificar que todas las medidas se realizan sobre la duración total del vídeo T_d , y que el usuario lo visualiza mediante la suma de un conjunto finito de I fragmentos de vídeo de duración c_i . Esto implica que un usuario u que consume íntegramente un vídeo de duración T_d cumple que:

$$\sum_{i=1}^I c_i = T_d \quad (3)$$

Con lo cual queda patente el tiempo total de consumo del vídeo por parte del usuario, y el número de paradas (I) que el usuario ha realizado durante la visualización.

Para un modelo clásico de consumo como el propuesto por Hyoseop sería suficiente con capturar la medida de duración de los tiempos c_i . Sin embargo, para el modelo propuesto en esta tesis es necesario conocer también los momentos en los que el consumo se inicia y se interrumpe. Por lo tanto, se establecen los conjuntos U y V , que determinan los instantes de inicio de un consumo normal (entendido como reproducción a la velocidad adecuada) y de cese de ese consumo normal (ya sea debido a una parada, *pause* o *stop*, o al inicio de otro tipo de actividad, como un *flash-forward*), respectivamente, medidos de forma relativa a la duración total del contenido. Por lo tanto, los valores de los conjuntos U y V caracterizan la reproducción normal del consumo audiovisual del usuario:

$$\begin{aligned} U &= \{u_1, u_2 \dots u_i\} \\ V &= \{v_1, v_2 \dots v_i\} \end{aligned} \quad (4)$$

Por ejemplo, en el caso de un consumo ininterrumpido de un contenido, de principio a fin, por parte de un usuario, $U=\{u_I\}$ y $V=\{v_I\}$, donde $u_I=0$ y $v_I=1$, dado que hacen referencia a los instantes relativos de inicio y fin del vídeo. Por lo tanto, el tiempo de consumo total del usuario u , T_u , viene dado por:

$$T_u = \sum_{i=1}^I (v_i - u_i) \quad (5)$$

Del mismo modo, de cara a otros posibles modelados del consumo, se establece también la medida de los instantes de inicio y de cese de reproducción del vídeo a una velocidad distinta de la normal, representadas por los valores de los conjuntos W y X

$$\begin{aligned} W &= \{w_1, w_2 \dots w_i\} \\ X &= \{x_1, x_2 \dots x_i\} \end{aligned} \quad (6)$$

La diversidad de dispositivos y de tipos de codificación de los vídeos impide una precisión mayor en la definición, ya que dependiendo de ambos factores pueden existir diferentes modos de reproducción, diferentes velocidades, etc. Por lo tanto, la caracterización más genérica y útil se obtiene mediante la captura de los valores U , V , W y X . La combinación de los diferentes valores de estos conjuntos permite advertir diversos comportamientos, como la reproducción normal, la pausa o parada, la reproducción a velocidad anormal, los saltos hacia adelante o hacia atrás, y la reproducción (ya sea a velocidad normal o anormal) en un sentido o en otro.

Como se verá en los próximos apartados, el estudio de estos valores será suficiente para la caracterización del interés de un usuario a partir de sus modos de consumo.

4.4 Modelo de transformación de consumo en interés

En vista de las necesidades expuestas en los apartados anteriores, se propone desarrollar un modelo estadístico capaz de calcular, en cualquier instante, el interés de un usuario por cualquier contenido o tipo de contenido consumido.

En primer lugar, es necesario comentar brevemente el corpus de aplicación sobre el que se diseña el modelo estadístico propuesto, así como sus características más relevantes.

El modelo estadístico propuesto pretende inferir el interés por un contenido audiovisual o por el tema de ese contenido a partir del consumo de contenidos en directo, es decir, aquellos que no están almacenados en un soporte físico (para lo cual, sin embargo, se podrían realizar adaptaciones). Es decir, se considera, por ejemplo, el consumo para una señal de difusión de televisión en un entorno *broadcast* clásico, o el consumo de vídeo en directo en *streaming* a través de Internet o, incluso, el consumo de contenidos ya almacenados en la Red, cuyo control pueda tener el usuario, es decir, que este pueda reproducirlos, pararlos, pasarlos en avance rápido, rebobinar, etcétera.

Tal y como se ha explicado en el apartado correspondiente al estado del arte, existen algunas aproximaciones al modelado de los consumos audiovisuales, pero, de todas ellas, sólo una sería directamente aplicable al caso del consumo audiovisual en directo que nos ocupa. Esta, además, es la aproximación más simple, consistente únicamente en una interpolación lineal del tiempo de consumo. Por ejemplo, si un usuario consume durante quince minutos un programa con una duración de treinta minutos, el interés inferido será del 50% de la escala (5 sobre 10, 2.5 sobre 5...).

Esta idea básica, además de suponer una aproximación muy gruesa que no considera ninguna variabilidad en el consumo audiovisual (en este caso, por ejemplo, el consumo de la primera mitad de un programa televisivo denotaría el mismo grado de interés que el consumo de la última mitad, cuando factores como los tiempos y el número de cambios podrían ser considerados), traslada únicamente el interés local por un contenido. Es decir, infiere un interés concreto sobre un contenido aislado, que no tuviera nada que ver con los demás.

Por otro lado, la aproximación basada en modelos de Markov está orientada a contenidos almacenados, en diferido, ya que el sistema se pone en marcha a partir de los eventos que el usuario aplica (reproducción, pausa, parada, avance rápido...), asociando

estos eventos, como observaciones, a los estados ocultos, que son los que denotan el grado de interés del usuario.

La tercera aproximación, que emplea inferencia bayesiana y árboles de decisión, no consiste estrictamente en una inferencia del interés, sino que se limita a operar sobre una base binaria, que considera un contenido consumido o no consumido, para pronosticar si se producirá o no un próximo consumo de otro contenido relacionado con el primero.

Partiendo, por lo tanto, de estas tres propuestas, se establece un modelo para el cual, en primer lugar, resulta importante diferenciar los siguientes conceptos:

- **Interés local.** Es el interés que provoca en el usuario el consumo de un contenido aislado, considerado independiente tanto a nivel temático como a nivel de una secuencia de contenidos enlazados. Su versión más básica sería la relación lineal entre el tiempo de consumo y el interés que se ha identificado en el estado del arte.
- **Interés global.** Se trata del interés que va acumulando un usuario a partir del repetido visionado de contenidos enlazados, es decir, aquellos que tienen algún vínculo de unión entre ellos. Algunos ejemplos de esto podrían ser:
 - El consumo de un telediario, o cualquier programa televisivo que se emita periódicamente. De cada emisión se obtendría un interés local, que serviría de dato de entrada para calcular el interés global que despertaría en un usuario, por ejemplo, el “telediario del canal N”. De esta manera, si un usuario por cualquier razón no puede ver el telediario un día concreto, esto no significaría que de repente el interés haya desaparecido. Esta es la razón que hace que el cálculo del interés global sea la base del sistema, contando con el interés local como dato de entrada.
 - El consumo de una serie de televisión. Para la emisión de cada capítulo se obtendría un interés local, que iría generando, a cada iteración, un interés global que dependería de él, y que así iría modificando moderadamente la percepción global de la serie que tuviera el usuario.
 - Varias películas de un mismo director. El interés local en cada película serviría para inferir el interés global del usuario en ese director, lo cual sería muy útil para una posterior recomendación que se basara también en contenido.
 - Contenidos cualesquiera de los que se pueda extraer una relación más o menos clara, pero que estaría modelada adecuadamente y permitiría extraer conclusiones particulares o generales acerca de las preferencias del usuario.

El estudio de las maneras de relacionar los diferentes contenidos a través del tema o de otras características de alto o bajo nivel, semánticas o sintácticas, es algo que debe realizarse de manera aplicada a cada uno de los escenarios de trabajo. En esta tesis se han contemplado dos casos bien diferenciados, aplicando el modelo a entornos categorizados (apartado 4.5) y a entornos multiatributo (apartado 4.6). Queda abierto como línea de trabajo futuro un estudio semántico más exhaustivo, que añada complejidad a las relaciones que se puedan extraer entre los distintos contenidos. Un eficiente modelado de estas relaciones podría reforzar de manera importante el sistema de inferencia de interés.

Volviendo al modelo propuesto, éste cumple las siguientes características:

- 1) Se considera que los contenidos no son aislados, por lo que una secuencia inicial de contenidos relacionados da lugar a una secuencia de consumos relacionados.
- 2) El sistema toma como entrada una serie de datos de consumo, que varían dependiendo del escenario de aplicación. Estos datos influyen en el número de factores a tener en cuenta de cara a la extracción del interés, como pueden ser factores de zapping, de compañía, etcétera, que se concretarán en los apartados de los modelos aplicados (4.5 y 4.6).
- 3) El sistema obtiene, en cada iteración, un valor de interés global con su correspondiente valor de precisión, cuya manera de obtenerlo figura en el apartado 4.4.2.

Por esta razón, en el modelo estadístico propuesto para el cálculo del interés se utilizarán en cierto sentido algunas ideas sugeridas del estudio del estado del arte.

En la Fig. 6 se puede apreciar una arquitectura de bloques general del conversor. Para la obtención del interés local, partiendo de la aproximación lineal clásica, se desarrollará (capítulos 4.5 y 4.6) en cada escenario concreto de aplicación un modelo que permita un ajuste más fino a la realidad. Por otro lado, también se tendrán en cuenta ciertos factores externos que influirán en el cálculo de la precisión y, por lo tanto, en la varianza de los sistemas iterativos diseñados en el módulo del interés global (general). Por lo tanto, son los bloques de interés local y factores de precisión, junto con sus adaptaciones (cuantificación y precisión de la inferencia) para el módulo final donde será necesario tratar las peculiaridades de cada tipo de escenario.

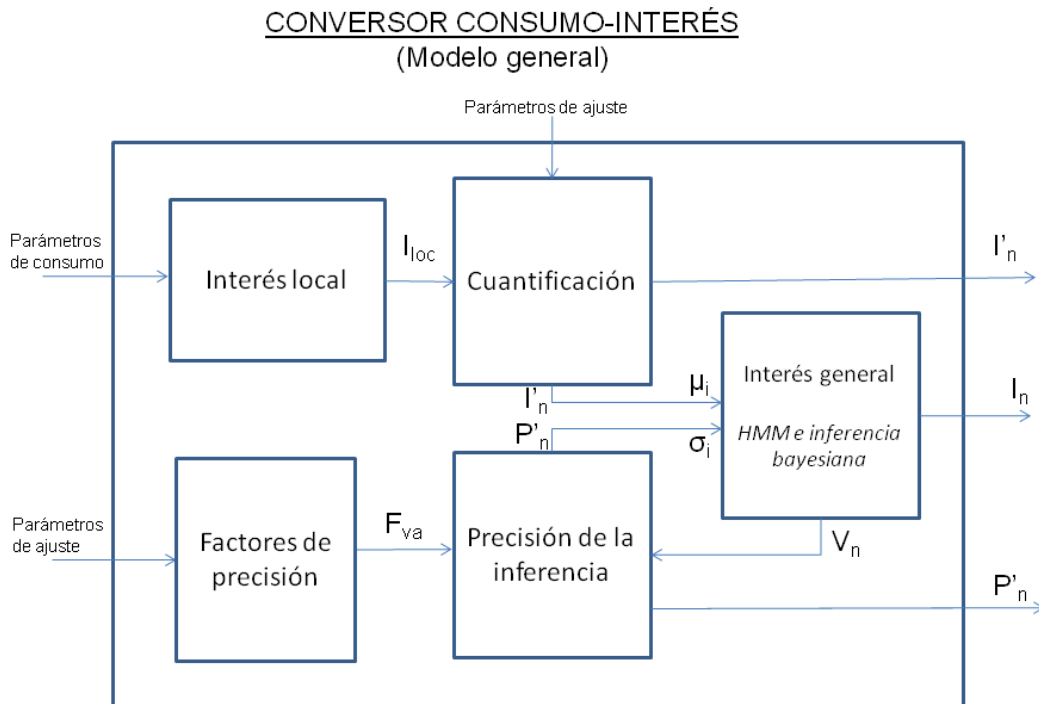


Fig. 6. Conversor consumo-interés. Modelo general

A partir de estos valores de salida de cuantificación y precisión, como se aprecia en la Fig. 6, se realizará, para obtener el interés global, un modelado estocástico de los consumos en forma de proceso oculto de Markov (HMM), con ciertas variaciones en su diseño y con el añadido de una etapa final de inferencia bayesiana.

4.4.1 Medida y cuantificación del consumo audiovisual (interés local)

Dentro del proceso global, en primer lugar se infiere el interés sobre un contenido aislado (interés local), es decir, emitido de manera puntual en un instante determinado. Posteriormente, se relacionará el dato resultante de esta deducción con la secuencia de intereses que se van infiriendo en cada iteración de la visión periódica. Este apartado explica la manera de obtener el valor del interés local.

La medida del interés local se presenta en una escala normalizada entre 0 y 1. Es decir, en el caso de que el consumo se produzca de manera íntegra de principio a fin, este valor debe ser 1, y en caso de que no haya consumo alguno, el valor será 0. En estos casos de valores extremos, los factores externos no aportan información alguna, por lo que desaparece su influencia en la obtención del resultado final.

Para el cálculo se tiene en cuenta el porcentaje de tiempo que el usuario consume sobre la duración total del contenido (100%), así como la distribución de este tiempo de consumo, dado que no es lo mismo consumir el inicio que el final de la emisión, ni consumir de forma continua que de manera más irregular.

Por ello, se consideran los instantes u_i , en los que el espectador inicia o reanuda la visualización del contenido, y los instantes v_i , en los que, por el contrario, el usuario deja de consumir el contenido, ya sea definitiva o temporalmente. Este es un detalle que será necesario tener en cuenta en una posible implementación posterior, a la hora de diseñar las características del sistema de captura de datos.

La expresión utilizada para inferir el interés local, que se puede ver en (7), se basa en la idea de que cada período de visualización del contenido (comprendido entre u_n y v_n) está ponderado por el instante en que se deja de consumir, v_n , dado que ese es el momento con un mayor valor para inferir el interés.

Se puede deducir entonces que el interés, en función de estas variables, viene dado por (7):

$$I_{LOC} = \sum_{n=1}^I (v_n - u_n) \cdot v_n \quad (7)$$

Donde:

I es el número de intervalos en que se ha producido consumo del contenido.

$$0 \leq u_n < 1$$

$$0 < v_n \leq 1$$

$$0 \leq I_{LOC} \leq 1$$

Para una mejor comprensión y visualización de los efectos, se aplica la fórmula a cuatro ejemplos. Las gráficas de Fig. 7, Fig. 8, Fig. 9 y Fig. 10 representan la secuencia de variación de consumo de un contenido aislado. El valor de ordenadas 1 corresponde a los instantes en que el espectador está visualizando el contenido, y el valor 0 se asocia a la suspensión de la visualización, ya sea por abandono o por consumo de un contenido diferente durante la emisión del contenido a estudio. Esta incertidumbre en la dicotomía entre suspensión o cambio de la visualización, será también tenida en cuenta posteriormente, en el cálculo de la precisión, a través del factor de zapping.

- a) Se considera que se consume un 30% del contenido según la distribución de la siguiente figura:

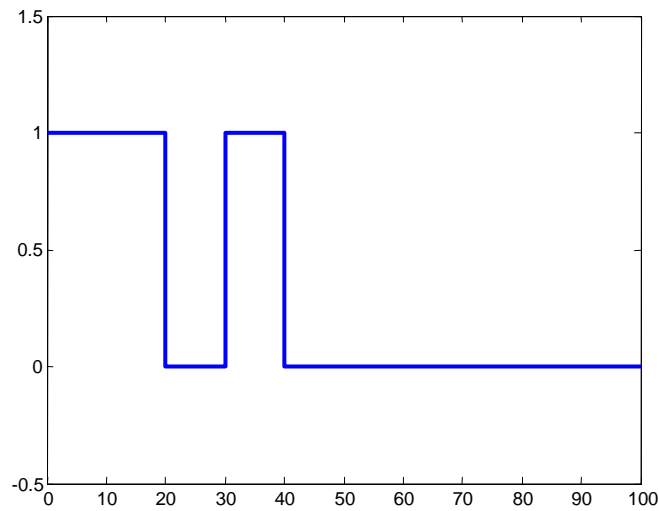


Fig. 7. Interés local. 30%. (a)

Los valores son $u_1=0$; $u_2=0,3$; $v_1=0,2$; $v_2=0,4$. Aplicando la fórmula de (7) resulta:

INTERÉS LOCAL= $I_{Loc}=0,08$

- b) Se considera que se consume un 60% del contenido según la distribución de la siguiente figura:

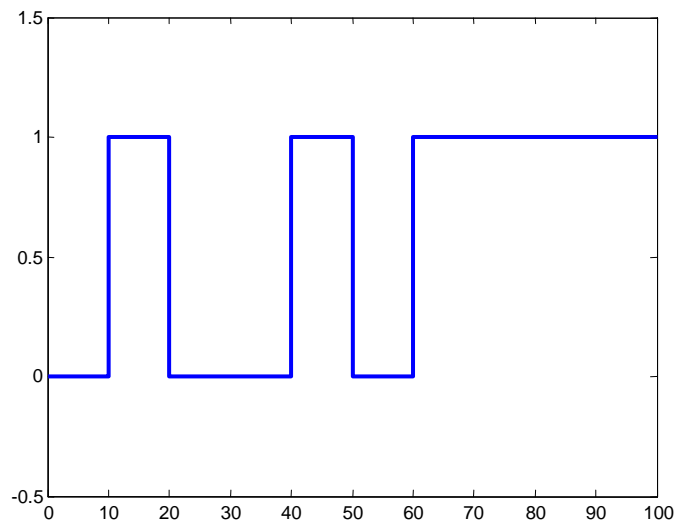


Fig. 8. Interés local. 60%. (b)

Los valores son $u_1=0,1$; $u_2=0,4$; $u_3=0,6$; $v_1=0,2$; $v_2=0,5$; $v_3=1$. Aplicando la fórmula de (7) resulta:

INTERÉS LOCAL= $I_{Loc}=0,47$

- c) Se considera que se consume un 60% del contenido según la distribución de la siguiente figura:

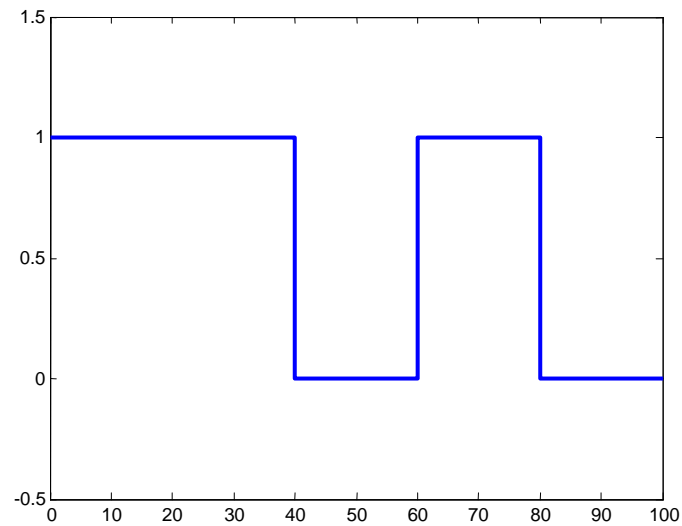


Fig. 9. Interés local. 60%. (c)

Los valores son $u_1=0$; $u_2=0,6$; $v_1=0,4$; $v_2=0,8$. Aplicando la fórmula de (7) resulta:

INTERÉS LOCAL= $I_{Loc}=0,32$

- d) Se considera que se consume un 30% del contenido según la distribución de la siguiente figura:

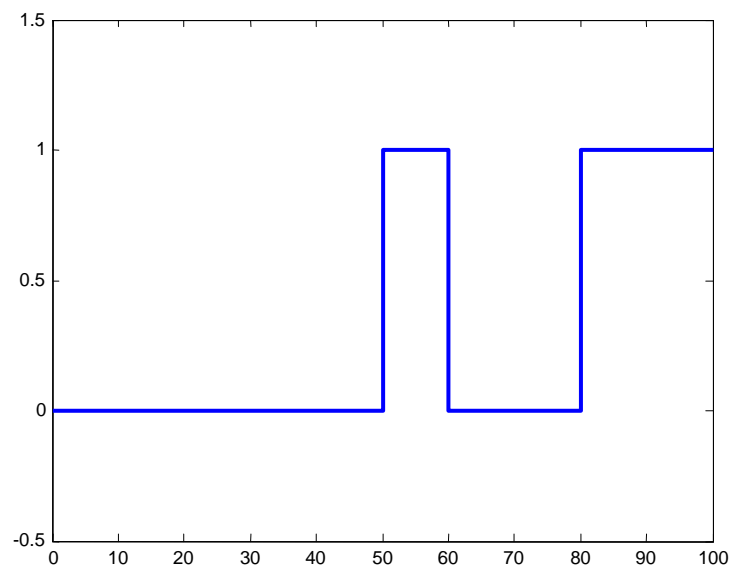


Fig. 10. Interés local. 30%. (d)

Los valores son $u_1=0,5$; $u_2=0,8$; $v_1=0,6$; $v_2=1$. Aplicando la fórmula de (7) resulta:

$$\text{INTERÉS LOCAL} = I_{\text{LOC}} = 0,26$$

Como se aprecia en los resultados de la Tabla 5, un mismo tiempo de consumo puede deparar diferentes resultados de interés, dependiendo de la distribución (la forma de realizar el consumo), con lo que la fórmula clásica lineal quedaría mucho más ajustada con esta nueva metodología.

	Prueba 1. Distribución balanceada al inicio	Prueba 2. Distribución balanceada al final
Tiempo de consumo = 30%	$I_{\text{LOC}}=0,08$	$I_{\text{LOC}}=0,26$
Tiempo de consumo = 60%	$I_{\text{LOC}}=0,32$	$I_{\text{LOC}}=0,47$

Tabla 5. Resultados de la aplicación del interés local

▪ Cuantificación

Los valores continuos obtenidos de interés local deben ser posteriormente cuantificados, de tal forma que los se correspondan con escalas discretas de valoración explícita, que permitan, además, la aplicación del algoritmo de obtención del interés global.

Por lo tanto, el número de niveles de cuantificación definidos debe coincidir con los niveles de la escala de preferencias con los que se diseñe el sistema⁹. Dado que se ha elegido un modelo de cinco grados de interés (que coincidirá posteriormente con el número de estados del Modelo Oculto de Markov), también se definirán cinco eventos diferentes de observación, que se obtienen de la cuantificación de la señal entrante de interés local.

Por este motivo, se definen cinco intervalos y niveles de cuantificación, que inicialmente se establecen de manera uniforme. Una posterior etapa de aprendizaje podría ir ajustando la anchura de los intervalos dinámicamente, según varíe el comportamiento de los usuarios y su manera de consumir puntualmente, es decir, contenidos aislados. De esta manera, la granularidad del cuantificador estaría regida por que el hábito de consumo del usuario esté más próximo o menos a una zona determinada, obteniendo así una mayor precisión.

Sin embargo, por simplicidad, dejando el desarrollo dinámico de la cuantificación como una inmediata línea futura, para la simulación implementada se toma modelo de cuantificador uniforme representado en la Fig. 11:

⁹ Esto no es imprescindible desde el punto de vista de la formulación matemática, pero permite que la salida de interés local y la salida de interés global del módulo de conversión de consumo en interés estén en la misma escala y, por lo tanto, sean comparables, por lo que se opta por esta aproximación. A esta ventaja, además, se añade una mayor sencillez a la hora de formular matemáticamente, en la siguiente etapa, el Modelo Oculto de Markov.

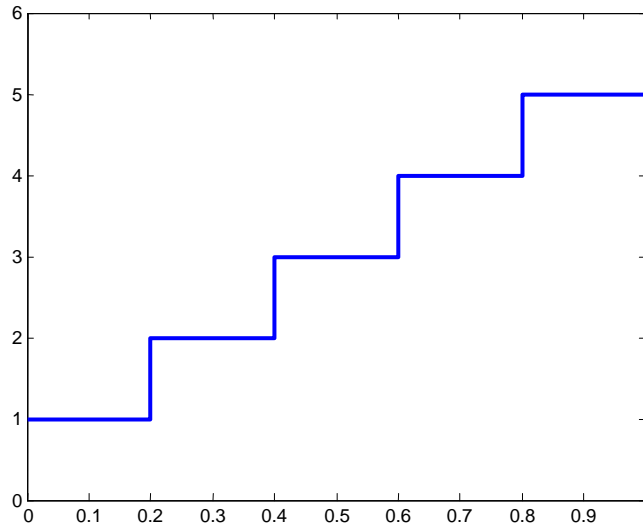


Fig. 11. Cuantificador uniforme empleado

De este modo, el cuantificador toma como entrada la salida del módulo de cálculo del interés local, cuyo rango está entre 0 y 1, y genera cinco posibles salidas distribuidas uniformemente, que coinciden con los diferentes eventos de observación posibles.

Para un número N de intervalos y valores de reconstrucción, el cuantificador uniforme seguiría la expresión indicada en (8):

$$I_c = I_{Loc} \cdot N \quad (8)$$

Es decir, el entero superior del producto del interés local de entrada por el número de intervalos del cuantificador.

4.4.2 Cálculo del interés acumulado del usuario (interés global)

La idea de utilizar como base un modelo oculto de Markov procede de la naturaleza de los datos de partida y de las características de lo que se quiere inferir. De entrada, se cuenta con una secuencia de consumos medidos y modelados como “interés local”. Estos serían los eventos observables, los cuales dependen del grado de interés que el usuario tiene en un determinado momento sobre un contenido. Este interés estaría cuantificado en diferentes grados, que representan los estados de la cadena de Markov. Estos estados (intereses) no son directamente medibles, sino que se deducen de los eventos observables, en este caso los datos cuantificados de interés local, que representan el nivel de consumo para un contenido aislado.

De esta manera, un modelo oculto de Markov (HMM) es el modelo matemático más ajustado para representar este proceso, según se aprecia en la Fig. 12:

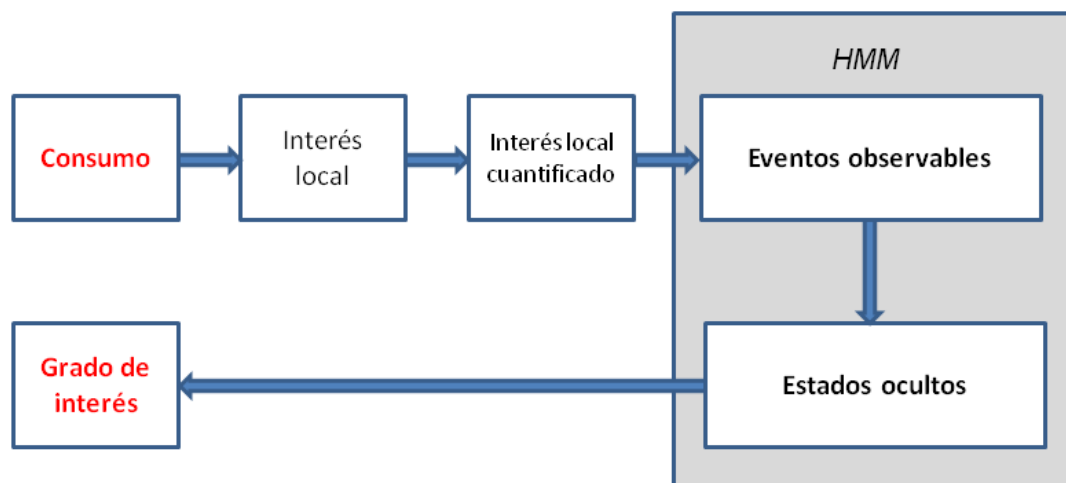


Fig. 12. Relación de la conversión del consumo en interés con un HMM

Siguiendo el esquema de la Fig. 12, se plantea un ejemplo para ilustrar el proceso. Un día concreto, el usuario del sistema ve la mayor parte del telediario de la noche de un canal de televisión. Sin embargo, un par de veces ha cambiado de canal y ha sintonizado otro programa. Sobre estos datos de consumo, el módulo de interés local extrae un valor que es inmediatamente cuantificado. Dado que el consumo ha correspondido a la mayor parte del telediario, pero no al contenido íntegro, el interés local cuantificado tiene un valor de 4 (sobre una escala que presenta un mínimo de 1 y un máximo de 5). Este valor 4 es uno de los posibles eventos observables, y supone un nuevo elemento para la cadena de eventos observables que alimenta al proceso oculto de Markov (HMM). El HMM, por lo tanto, tiene la misión de inferir el estado oculto que corresponde a este último evento, y que se corresponderá con el grado de interés que el usuario tiene en ese momento sobre ese telediario, y que cada día se va modificando según la variación de los consumos. Si, por ejemplo, el usuario acostumbra a ver el telediario íntegramente, el hecho de que en esta ocasión el consumo haya sido inferior, seguramente no provoque un descenso del grado de interés de 5 a 4, ya que la causa del menor consumo puede estar en múltiples circunstancias externas, pero si este consumo sigue la línea decreciente, los efectos se notarán en el cálculo del interés, que terminará descendiendo.

Volviendo a la explicación teórica, es preciso recordar que el modelo diseñado no se puede basar en un proceso oculto de Markov clásico, debido a que una serie de peculiaridades obligan a realizar una serie de adaptaciones. Las diferencias del modelo propuesto con un HMM son las siguientes:

- En un modelo oculto de Markov, un suceso para el cual se establece una observación lleva implícito el estado que se pretende determinar. En este caso, el consumo en un instante n , que normalmente sería lanzado por el estado en que se encuadra el sistema en el momento n , estará provocado por el interés (estado) en el instante $n-1$. Esto se debe a que un usuario consume un contenido en un instante según el interés que se estableció en el consumo anterior. De esta manera, el interés se presenta como una consecuencia del consumo actual, el cual depende del interés anterior. Por lo tanto, el uso de un HMM clásico llevaría a obtener un estado del interés en la iteración anterior al momento en estudio.

- Otro factor que influye en la variación del HMM es la imposibilidad de realizar un aprendizaje del acierto de los estados de manera objetiva. Es decir, no se puede cuantificar mediante una metodología intrínseca que el interés inferido es el correcto. Esto lleva a que los estados de interés se presenten de manera probabilística, evitando cualquier conjetura determinista.

El nuevo modelo, por lo tanto, seguiría el esquema mostrado en la Fig. 13, en el que las flechas en azul corresponden a las probabilidades de transición entre estados (línea continua) y entre observaciones y estados (línea discontinua), mientras que las flechas rojas representan una probabilidad que debe ser obtenida por otros métodos (en el modelo propuesto, mediante la etapa final de inferencia bayesiana). La Fig. 13 se presenta de una manera simplificada por legibilidad y sencillez, sin considerar todas las transiciones posibles, pero es necesario tener en cuenta que se trata de un sistema ergódico, ya que desde cualquiera de los estados definidos es posible pasar a cualquier otro estado.

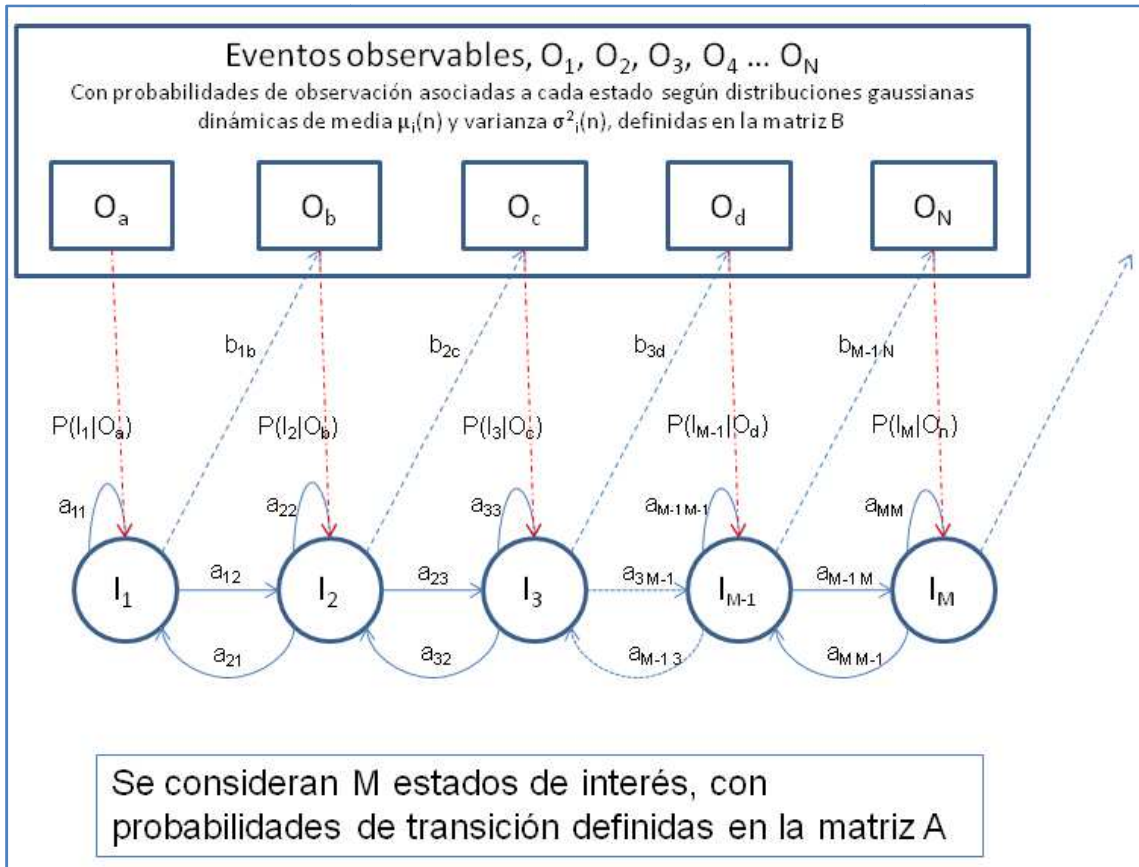


Fig. 13. Adaptación de un HMM al problema de inferencia del interés

En este punto, para la obtención del estado (grado de interés) a partir del consumo actual (flechas rojas en la Fig. 13), se han utilizado técnicas de inferencia bayesiana que también contemplen la probabilidad de transición entre estados.

Definición del modelo oculto de Markov (HMM) adaptativo

El modelo oculto de Markov está definido por los tres parámetros de (9).

(9)

Donde π es el vector de probabilidades iniciales, que establece las probabilidades del estado inicial del sistema; A es la matriz de probabilidades de transición entre estados, y B es la matriz de probabilidades de observación.

Se define A , la matriz de probabilidades de transición entre estados, según (10), como una matriz cuadrada de dimensión $m \times m$, donde m es el número de estados (grados de interés) considerados.

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mm} \end{pmatrix} \quad (10)$$

Para la creación de esta matriz, sabemos que $\sum_{j=1}^m a_{ij} = 1$, es decir, que la suma de cada fila debe ser 1, la probabilidad total, ya que representa la probabilidad existente, en cada uno de los estados, de llegar en el siguiente paso a cualquiera de los estados posibles. Por definición, a_{ij} representa la probabilidad de que haya una transición del estado i al estado j .

Por lo tanto, el cálculo de la matriz A y las posteriores actualizaciones con cada iteración se va a gestionar fila a fila, generando y modificando las distribuciones estadísticas establecidas.

Concretamente, se considera cada una de las filas de A como una distribución normal, cuya media y varianza parten de un estado inicial y se van actualizando con cada iteración, según el sistema va descubriendo la “facilidad” del usuario para cambiar su estado de interés relacionado con el tema o contenido en estudio. Para este caso, no se van a considerar parámetros generales del usuario, que puedan ser deducidos a partir de otros consumos independientes, ya que ese factor ya se tendrá en cuenta para el cálculo de la matriz de observaciones B . Sin embargo, también sirve la matriz A para la regulación de los consumos según la precisión de entrada, procedimiento que se explicará en el punto de este mismo apartado “Medida de validez y medida de precisión”.

Se define, para cada fila de la matriz A , el estado inicial mostrado en (11).

$$A_i \text{ inicial} \begin{cases} \text{Distribución gaussiana} \\ \text{Media: } \mu = i \\ \text{Varianza: } \sigma^2 = 1 \end{cases} \quad (11)$$

Se toma la varianza inicial de la distribución gaussiana estándar, ya que resulta la más neutral para el instante en que aún se carece de información del usuario. Del mismo modo, la distribución toma como valor medio el estado anterior (que corresponde al número de fila de la matriz), debido a la denominada, “resistencia al cambio”, habitual en los patrones de consumo humano, ya sean estos audiovisuales o de cualquier otro tipo.

Para la actualización de A , se tiene en cuenta la nueva información que se va obteniendo acerca de la facilidad o dificultad del usuario para cambiar de estado o permanecer en el

mismo. Por lo tanto, se redefine el valor de la media para cada fila de manera iterativa, según (12).

$$\mu_n = \mu_{n-1} + c(i) \cdot \frac{I_n - \mu_{n-1}}{n} \quad (12)$$

Donde $c(i)$ es el coeficiente de variación de la media, que expresa la variación del grado de incertidumbre en el cambio de estado por parte del usuario, por lo que cambiaría según la fiabilidad de los datos almacenados de cada usuario. Por otro lado, el valor de la varianza, a pesar de expresar la propensión del usuario a cambiar de estado, y dado que esto queda incorporado indirectamente en la fórmula de actualización de la media, servirá para modular en cada iteración la influencia de cada observación dependiendo de los factores externos, tal y como se explica en el apartado de la medida de validez. En el caso de no existir factores externos, su valor permanecerá en el inicial por defecto.

Por último, la matriz B , de observaciones, puede presentar un tamaño de $m \times n$, siendo m el número de estados y n el número de observaciones, pero en esta descripción, como ya se ha explicado, se considera un número de observaciones igual al número de estados, por lo que se tendría una matriz $m \times m$ como la representada en (13).

$$B = \begin{pmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mm} \end{pmatrix} \quad (13)$$

Donde b_{ij} representa la probabilidad de que se alcance el estado i a partir de una observación j . Esta definición sería la inversa de un HMM clásico, debido a la asincronía temporal que ya se ha comentado.

Su modelado se presenta de manera análoga a la matriz A , consistiendo en una distribución gaussiana en cada una de sus filas, modelándose inicialmente como muestra (14).

$$B_i \text{ inicial} \begin{cases} \text{Distribución gaussiana} \\ \text{Media: } \mu = i \\ \text{Varianza: } \sigma^2 = 1 \end{cases} \quad (14)$$

Es decir, se utiliza de inicio el valor estándar de la varianza y se centra la distribución en el elemento de $i=j$ de la fila i , al tratarse de una matriz cuadrada, correspondiente así a la probabilidad de observación del evento equivalente al estado actual.

En este caso, para la actualización de B sólo se modifica la varianza, permaneciendo la media constante en el valor indicado, siempre que se cumpla la premisa de realizar una cuantificación de intereses locales en un número de niveles igual al número de estados de interés del sistema.

Pero antes de proceder a la explicación de la forma de actualizar la matriz B , es necesario comentar la naturaleza y estructura de esta propia matriz.

El valor de la probabilidad de que se tenga un interés i después de haber realizado un consumo o se puede descomponer en el producto de dos coeficientes: b' y γ . B' , en cada una de sus filas, expresa la dispersión de la relación del consumo periódico con el

interés. Esto es, indica la propensión que un usuario tiene a consumir un determinado contenido según el estado en que se encuentre. Por ejemplo, un usuario puede tener un gran interés en un determinado contenido sin que esto implique que lo consuma regularmente, lo cual se podría inferir, llegado el caso, si la secuencia de consumos fuera muy divergente (presentara un alto valor medio de frecuencias). Sin embargo, si los consumos tienden a converger para un determinado estado, significa que el usuario presenta una respuesta muy regular ante el interés, ya que, llegado el caso, siempre vería un programa que le gusta, o nunca volvería a consumir uno que le ha disgustado. Denominamos a esta función matricial, B' , como la **divergencia espacial de consumos**. El otro coeficiente, γ , expresa la subjetividad y **efusividad del usuario en relación a los consumos**. Se trata, a diferencia de B' , estructurado según dos dimensiones, interés y observación, de una variable dependiente de una única dimensión, por lo que se puede expresar mediante un vector, que expresa evolución de la efusividad en función del estado.

Por lo tanto, quedan resumidas las ideas principales en el esquema de la Fig. 14:

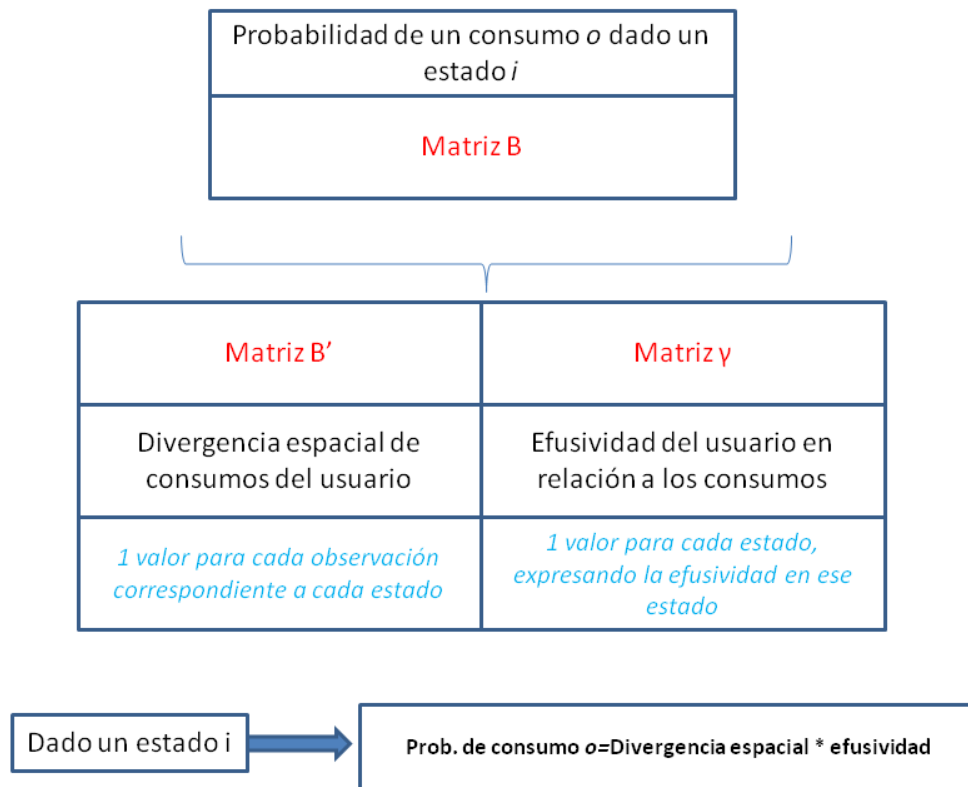


Fig. 14. Relación de la matriz B con sus submatrices

Formalizado matemáticamente, se presenta la relación de la matriz B con B' y γ en (15):

$$B = \begin{pmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mm} \end{pmatrix} = \begin{pmatrix} b'_{11} \cdot \gamma_1 & \cdots & b'_{1m} \cdot \gamma_m \\ \vdots & \ddots & \vdots \\ b'_{m1} \cdot \gamma_1 & \cdots & b'_{mm} \cdot \gamma_m \end{pmatrix} = \begin{pmatrix} b'_{11} & \cdots & b'_{1m} \\ \vdots & \ddots & \vdots \\ b'_{m1} & \cdots & b'_{mm} \end{pmatrix} \cdot \begin{pmatrix} \gamma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \gamma_m \end{pmatrix} \quad (15)$$

Como se puede ver, γ sólo aporta un factor de ponderación y, aunque sea un vector, se expresa como una matriz bidimensional para poder operar directamente sobre B' y formar la matriz B.

De este modo, el valor del vector γ se calcula desde fuera del sistema, dado que refleja un carácter general del usuario, su comportamiento de efusividad, para lo que se emplean factores externos y datos del perfilado de usuario que deben ser modelados aprovechando información más heterogénea. En este modelado se va a considerar constante, es decir, una distribución uniforme, de manera que no influya y el peso real de la matriz B esté en la obtención de B'. Se realiza esta aproximación debido a que la matriz A ya recoge los valores generales de efusividad del usuario, asociados a un determinado grado de interés.

Por lo tanto, el valor inicial de B' coincide con el ya indicado para B: una distribución gaussiana de media μ y varianza la unidad.

Conceptualmente, se resume en la Tabla 6 el significado de los estadísticos asociados a las distribuciones de las filas de cada una de las dos matrices que configuran el HMM.

	Efusividad	Resistencia al cambio
General	Media matriz A	Varianza matriz A ¹⁰
Cada ítem global	Media matriz B	Varianza matriz B

Tabla 6. Caracterización de las distribuciones matriciales

Por lo tanto, la matriz A, de transición entre estados, está relacionada con el comportamiento general del usuario, y no solo con un determinado ítem. Por esta razón, el sistema de recomendación debe compartir los parámetros estadísticos de A para todos los ítems globales que tenga definidos. Por su parte, los valores de B son específicos para cada ítem global. En cuanto al significado de los dos estadísticos que definen las distribuciones gaussianas, la media se emplea para modelar la diferente efusividad de los usuarios, ya que para un mismo grado de interés, dos usuarios distintos pueden consumir cantidades muy diferentes de un contenido audiovisual; por otro lado, la resistencia al cambio de cada usuario está graduada por la varianza, ya que una mayor anchura de la distribución implica que el usuario es más voluble a cambiar sus hábitos de consumo y, por ejemplo, un usuario puede presentar grados de interés más altos rápidamente, después de menos consumos que otro usuario con mayor resistencia al cambio, mayor persistencia de hábitos.

En cuanto a la actualización de las distribuciones, análogamente a A, se considera que la media se mantiene constante en su valor inicial. Será la varianza, por lo tanto, el dato que aporte la información dinámica a la estructura de la matriz B. Para obtener el nuevo valor de la varianza con cada iteración, hay que estudiar la variabilidad de la secuencia de entrada de consumos observados, es decir, el factor de variación de la divergencia espacial, entendido el espacio como el recorrido con consumos posibles para un estado determinado.

Para obtener este nuevo valor de la varianza se recurre al análisis espectral de la secuencia de consumos con que se cuenta como datos. En primer lugar, como cada varianza corresponde a un determinado estado, se forma para cada estado un vector (a partir del vector inicial con la secuencia de todos los consumos del contenido

¹⁰ Este valor también incluye la fiabilidad de la iteración dada por los factores externos, dado que el concepto de resistencia al cambio puede ser absorbido en la actualización de la media.

enlazado¹¹) integrado únicamente por aquellos valores de consumo cuantificado de la secuencia inicial que se hayan registrado en el instante n, siendo ese estado el actual en el instante n-1. Es decir, se agrupan los consumos que hayan sido consecuencia de un determinado grado de interés.

Sobre este nuevo vector, es necesario obtener el grado de variabilidad de los datos, ponderando no sólo el número de variaciones, sino también la manera en que se producen esas variaciones, cuánto es el cambio que se produce. Para ello, lo más adecuado resulta un análisis de la secuencia en el dominio de la frecuencia, sobre la cual se obtiene el valor de la potencia media después de haber quitado la componente continua. Es decir, se realiza una transformada discreta de Fourier (DFT) por medio de alguno de los algoritmos habituales (FFT, DCT...), se elimina la componente continua, y se extrae de la señal resultante el valor de la potencia media, que, multiplicado por un coeficiente que influirá en el grado de suavizado del sistema global, dará directamente el valor de la nueva varianza de la matriz B'. Se aprecia el proceso gráficamente en la Fig. 15.

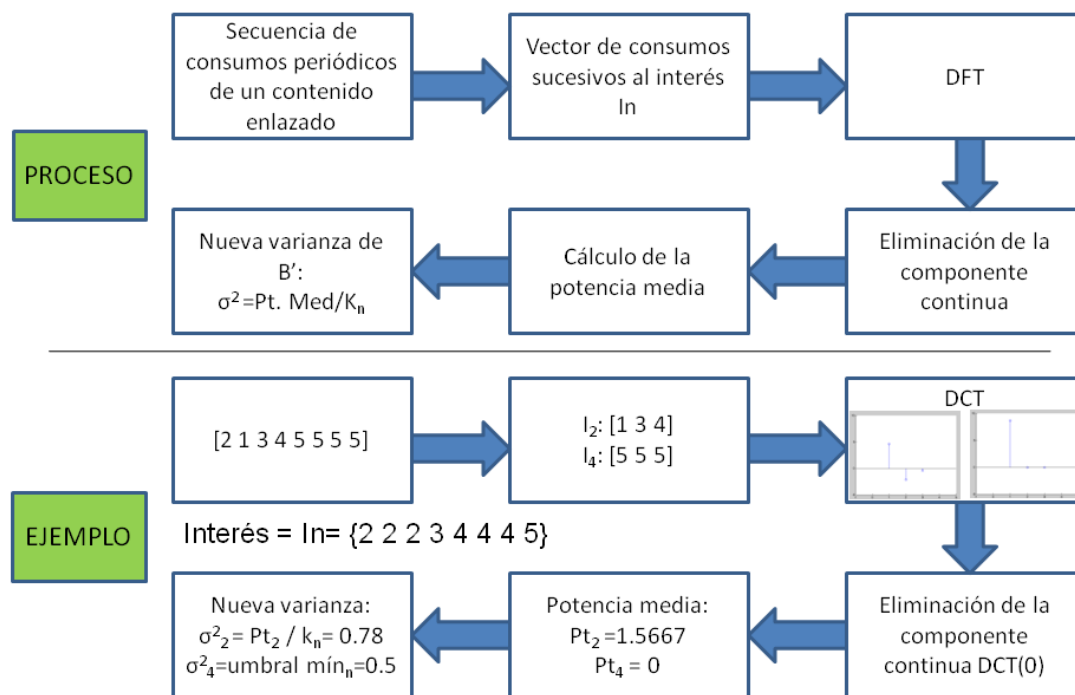


Fig. 15. Proceso y ejemplo de la variación de la matriz B

Como se advierte en la Fig. 15, es necesario establecer un umbral mínimo de dispersión, ya que, especialmente en las primeras iteraciones, es muy probable que se sucedan consumos cuantificados iguales, lo que llevaría la nueva varianza a un valor 0 que anclaría el sistema y produciría un único valor fijo del sistema, que correspondería a la media. Por lo tanto, queda en manos del diseñador fijar un umbral que considere adecuado; para las pruebas realizadas en esta tesis, ese valor fue fijado en 0,5.

Finalmente, es importante tener en cuenta que las metodologías de actualización de las matrices del HMM son sólo sugerencias del propio modelo general, pero la decisión de la estrategia específica está condicionada al escenario de aplicación y a las posibilidades

¹¹ Aquellos contenidos cuya emisión se repite periódicamente

y requisitos del software en el que se implemente. Existen otras muchas maneras, por ejemplo, de calcular la dispersión de consumos por estado que luego se asocia a la varianza de la matriz B, y será el diseñador quien tome la decisión en función del escenario.

Cálculo final del interés global

Con los valores de probabilidad calculados y actualizados en cada matriz, es posible realizar un cálculo final del interés global que un usuario tiene, en un momento determinado, por el tema de un contenido, por el propio contenido si se trata de contenido enlazado, etcétera. Para ello, se utiliza una sencilla técnica de inferencia bayesiana.

La probabilidad final que se desea calcular es la probabilidad estar en el estado de interés $n+1$, teniendo en cuenta que se ha realizado la observación de consumo cuantificado O_{n+1} y que el estado de interés en el instante n era I_n . Se consideran ambas probabilidades de transición como sucesos independientes, por lo que se puede expresar de la forma:

$$P_n(I_{n+1}|I_n, O_{n+1}) = P(I_{n+1}|I_n) \cdot P(I_{n+1}|O_{n+1}) \cdot P(I_{n+1}) \quad (16)$$

Aplicando el teorema de Bayes al segundo término del segundo miembro resulta:

$$P_n(I_{n+1}|I_n, O_{n+1}) = P(I_{n+1}|I_n) \cdot \frac{P(O_{n+1}|I_{n+1}) \cdot P^2(I_{n+1})}{P(O_{n+1})} \quad (17)$$

Y desarrollando:

$$P_n(I_{n+1}|I_n, O_{n+1}) = P(I_{n+1}|I_n) \cdot \frac{P(O_{n+1}|I_{n+1}) \cdot P(I_{n+1}|I_n) \cdot P(I_n)}{P(O_{n+1})} = a_{n,n+1} \cdot \frac{b_{n+1,n+1} \cdot a_{n,n+1} \cdot \widehat{P}_{n+1}}{1} \quad (18)$$

Donde \widehat{P}_n es el valor de la fiabilidad del estado anterior, es decir, el candidato de probabilidad del interés global elegido en el estado anterior por ser el de mayor valor.

Con esto, se calcula la probabilidad de estar en cada uno de los estados posibles y, entre todos ellos, el valor máximo será elegido, correspondiendo I_n al interés final inferido. El valor de esta probabilidad, asimismo, se utiliza como dato de salida, expresando la fiabilidad de todo el proceso. Esta fiabilidad puede verse realimentada por más parámetros, cuya influencia queda reservada a cada escenario concreto de aplicación, dos de cuyos casos se desarrollarán en los próximos apartados.

Finalmente, a modo de resumen, se presenta en la Fig. 16 un diagrama de actividad del proceso de inferencia del interés global a partir del interés local, en el que están involucrado tanto el HMM como la etapa de inferencia bayesiana.

Medida de validez y medida de precisión

Con el fin de dotar de consistencia a este valor del interés global, se introducen ciertas medidas de fiabilidad de la inferencia realizada. Para ello, se introducen dos parámetros, que se definen como “medida de validez” y “medida de precisión”.

Estas medidas pretenden dar una idea, por un lado, de la utilidad de cada medida de consumo individual a la hora de introducir el nuevo dato en la secuencia de consumo global (medida de validez) y, por otro, de la probabilidad de acierto del interés final calculado según el número de datos con que se cuenta (medida de precisión). Por este motivo, la medida de validez presenta un valor independiente de rango absoluto (entre 0 y 1) para cada iteración del sistema (cada vez que se produce un consumo), mientras que la medida de precisión es acumulativa, dependerá de la de validez, y siempre tendrá un valor igual o superior al presentado en el instante anterior.

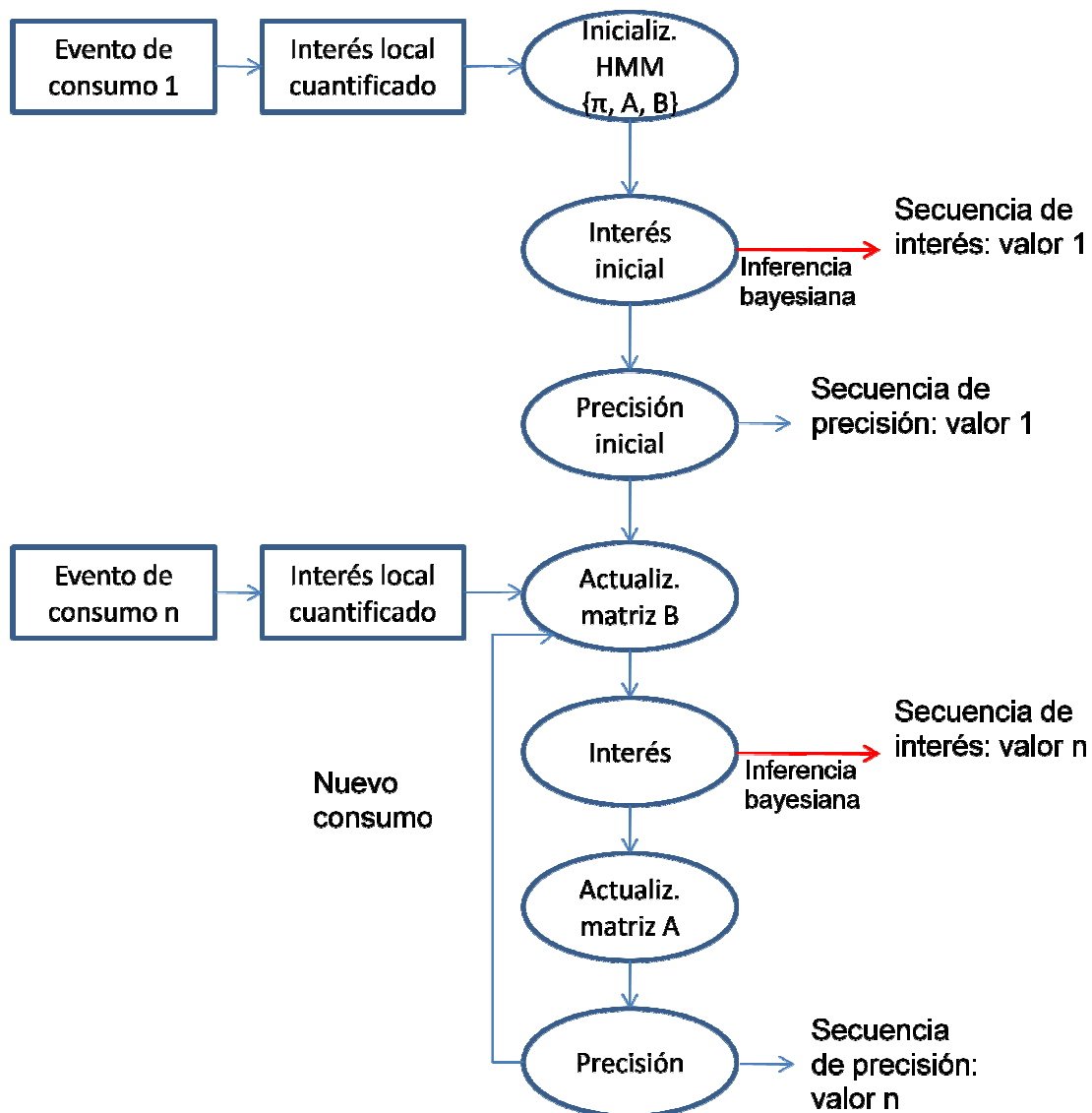


Fig. 16. Algoritmo de obtención del interés global

Pero esta precisión no sólo influye en las medidas de salida descritas, sino también en la propia medida final del interés inferido. Esto es, si por ejemplo se produce una medida de consumo que lleve asociada una precisión muy baja (debido a los factores externos de precisión, como puedan ser el factor de compañía o el factor de zapping), esa medida no debe ser apenas considerada para inferir el nuevo estado de interés. De esta manera, se concluye que estos factores actúan directamente sobre el valor de la varianza asociada al estado actual de la matriz A.

La matriz A, como se ha especificado en (10) y (11), indica la probabilidad del cambio de estado (pasar de un interés a otro), por lo que si un consumo no es representativo del interés del usuario (según la indicación de estos factores ya comentados), este no debería influir en el grado de interés inferido, y el estado debería mantenerse. Por esta razón, si un individuo, por ejemplo, se encuentra en el estado de interés 3, y llega el interés local de un consumo producido con precisión nula, la fila 3 de la matriz A se reconstruirá, manteniendo la media, con una varianza 0, con lo que no se producirá cambio alguno de estado. Análogamente, para valores superiores de precisión, el cambio de estado se podrá considerar de manera más fiel, por lo que hará aumentar el valor de la varianza para el estado en cuestión.

Finalmente, se concluye que la validez se obtiene como el producto de la probabilidad proveniente de la etapa de inferencia bayesiana por la precisión local de la medida de consumo. La precisión de salida, por su parte, dependerá de la evolución de esta secuencia de validez, ya que como se ha comentado es acumulativa, y la fórmula concreta se implementará para cada escenario de aplicación según esta idea de base.

4.5 Aplicación a entornos categorizados

Una vez desarrollado el modelo general, es necesario aplicarlo a un escenario concreto, de manera que queden delimitados los posibles datos de entrada del sistema y las relaciones entre ítems.

Se considera aquí un escenario *broadcast*, en el que un usuario recibe los contenidos, por ejemplo, a través de un receptor de televisión digital terrestre, sobre el que existe un software de captura de datos de usuario, consumo y contexto, que son recogidos por el módulo conversor de consumo a interés con el objetivo de que otro módulo realice la recomendación de contenidos final.

De cara a la aplicación del modelo general, resulta importante identificar en este escenario tanto los ítems locales como los ítems globales. Se define como ítem local cada programa emitido con una duración continua y compacta. Los ítems globales, por su parte, se definen como las categorías en las que se pueden agrupar los diferentes ítems locales, o aquellos programas con una emisión periódica, es decir, que emiten con una frecuencia diaria, semanal, etc. Cada uno de los ítems locales pertenece a un ítem global con un determinado peso, que depende de su correlación con la categoría a la que pertenece. Por ejemplo, una determinada serie de televisión puede ser considerada un ítem global, siendo cada uno de sus episodios un ítem local. En este caso, todos los ítems locales estarían al mismo nivel, y contribuyen con un peso 1, que se considera máxima correlación. Otro ejemplo similar es el telediario emitido por una cadena de televisión diariamente en una determinada franja horaria, constituyendo un ítem global. Cada día, sus emisiones diarias corresponden con los ítems locales, y el modelado del interés para cada uno de estos ítems se irá acumulando paulatinamente y realimentando el valor del ítem global “Noticias en el canal N”.

Este es el caso más sencillo, pero también es posible realizar una definición más compleja de ítems globales y de los ítems locales que los configuran con diferentes pesos. Por ejemplo, en un ítem global que fuera “literatura del siglo XX”, podrían contribuir con diferentes pesos distintas piezas audiovisuales, como una tertulia

literaria, un documental, una película sobre un escritor contemporáneo, o una noticia sobre el tema incluida en un telediario.

Esta flexibilidad permite aplicar los conceptos a los más diversos escenarios, adaptados a distintos propósitos comerciales. Sin embargo, en la aplicación desarrollada en esta tesis para la validación del modelo solo se han considerado pesos de valor 1 de los ítems locales, de tal forma que un ítem local solo pueda pertenecer a un ítem global, correspondiente con la categoría principal que se asocia a un determinado programa televisivo, según clasificaciones televisivas estándar, como la propuesta por TV-Anytime [23]. El apartado de pruebas define concretamente los ítems globales elegidos para la validación del modelo.

Una vez definidos los ítems de los que el sistema debe extraer el interés que despiertan en los usuarios, se puede continuar describiendo la adaptación del modelo general al caso de uso de transmisiones *broadcast* de televisión tradicional.

Siguiendo la nomenclatura de los modelos ocultos de Markov (HMM) ya definidos anteriormente, se concretan en el modelo adaptado 5 estados de interés, que son inferidos por el sistema en el bloque “Evolución del interés general”, especificado en la Fig. 17. Como se aprecia en la mencionada figura, las tres salidas del sistema son el interés local cuantificado del ítem (que puede resultar útil como complemento al interés anterior de cara a la recomendación), el interés global asociado a su categoría, y el valor total de precisión.

En este escenario se definen dos factores externos, el factor de zapping y el factor de compañía, que reciben como entrada una serie de parámetros de ajuste, además del dato fundamental de entrada: en un caso, el número de veces que se produce cambio de canal durante la emisión del contenido en estudio y, en el otro, el número de personas que, en ese momento, están viendo la televisión.

El esquema de la Fig. 17 muestra la aplicación general del módulo conversor de consumo a interés a un escenario *broadcast*.

Además de los factores externos comentados, el punto más importante de la adaptación a un escenario concreto está en señalar la naturaleza de los datos de entrada al módulo de cálculo del interés local.

El sistema de captura de datos, insertado en el set-top-box de cada usuario, resultaría análogo a un sistema de medición de audiencias, por lo que se considera que se cuenta como datos de entrada, para cada consumo, con una recopilación de los instantes de tiempo en que se ha comenzado o dejado de ver el contenido en cuestión. El vector v de las figuras Fig. 7, Fig. 8, Fig. 9 y Fig. 10 representa los instantes de tiempo en los que se deja de ver el contenido y, el vector u , los instantes en que se empieza o reanuda cada visualización.

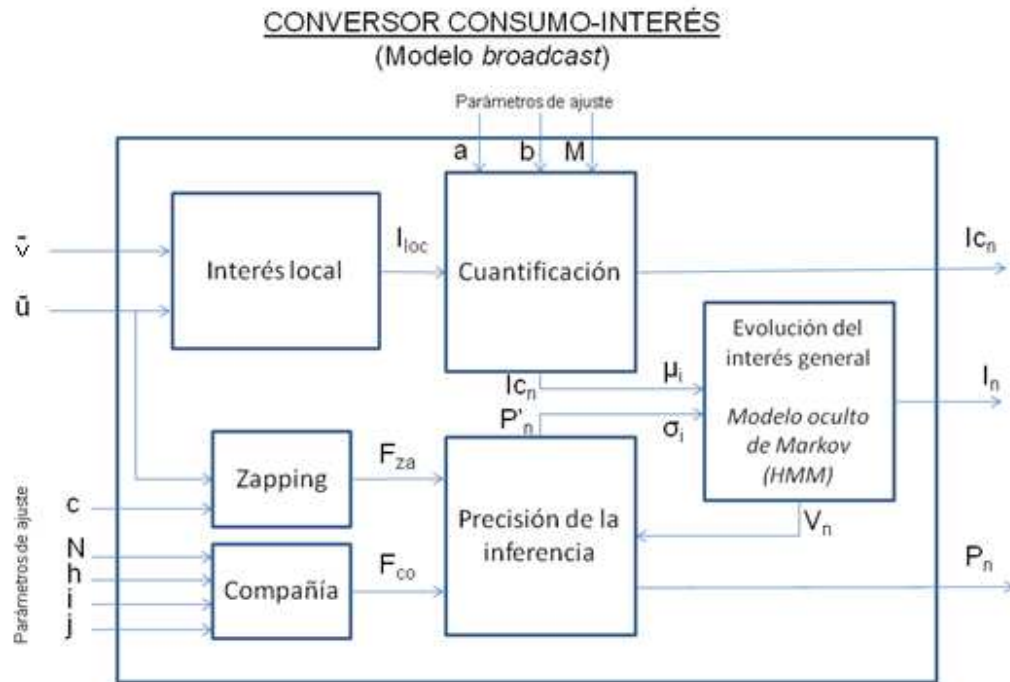


Fig. 17. Conversor consumo-interés. Modelo *broadcast*

Por esta razón, desde ese punto de vista, las dos entradas v y u serían vectores que, por sí solos, únicamente tendrían sentido para el módulo de interés local. Debido a esto, se considerará que los vectores v y u forman parte de una matriz en la que dos dimensiones representan esos instantes de tiempo frente a la dicotomía “ver”-“dejar de ver”. La tercera dimensión agrupa estos vectores en matrices, cada uno de cuyos elementos representa un contenido enlazado. Todos ellos juntos, es decir, la matriz tridimensional, muestra la manera en la que se han consumido todos los contenidos de una misma serie lógica de elementos (series de televisión, contenidos relacionados por un mismo tema, etcétera). A su vez, estas cadenas de consumos enlazados pueden unirse, en una mayor capa de abstracción, a nivel de usuario, para obtener nueva información que se utilizaría, como se ha definido en el apartado anterior, a través de la matriz B especificada en (15). Para el desarrollo de la implementación y las simulaciones, se ha elegido el nivel de abstracción inmediatamente inferior, es decir, sólo se ha considerado una secuencia de contenidos enlazados de un mismo usuario, en la que, además, también aparece la agrupación de secuencias enlazadas de ese usuario.

Factor de zapping

Un factor que influye en la fiabilidad del resultado del grado de interés obtenido es el zapping, que en un escenario de *broadcast* sirve como dato de entrada para la obtención de la precisión del cálculo. Un mayor zapping por parte de un usuario debe reducir la fiabilidad de las medidas obtenidas, ya que denota:

- a) Aunque el usuario esté consumiendo mayoritariamente un contenido, hace zapping porque busca algo que le satisfaga más, por lo que ese consumo mayoritario puede no deberse a un alto interés, sino a la ausencia de otros

- contenidos interesantes. Un mayor número de cambios de canal implica mayor incertidumbre en la correlación entre consumo e interés.
- b) Que el usuario presente un consumo pequeño sobre un contenido, en presencia de zapping, no significa necesariamente que ese contenido no le satisfaga, ya que puede haber otro contenido simultáneamente en otra cadena de televisión que le atraiga en un grado similar. En ese caso, ambos contenidos le atraen mucho, aunque el consumo no lo denote, por lo que es preciso disminuir la fiabilidad de la medida, que sería inversamente proporcional al número de cambios de canal.

Por lo tanto, estima la ponderación a través de (19):

$$F_{za} = 1 - \frac{c}{N} \quad (19)$$

Donde $N > 0$ es el número de veces que se ha cambiado de canal, y c es un parámetro que se debe ajustar a partir de los datos reales de consumo del usuario.

Esta función no está definida para un valor de $N=0$, es decir, para el caso en que no se haya producido ningún cambio de canal, en cuyo caso $F_{za} = 1$.

Factor de compañía

El otro parámetro que modela la validez de la medida es el factor de compañía. Cuanta más gente esté consumiendo un contenido audiovisual simultáneamente menos probable es que el consumo se deba a la voluntad de cada individuo. Sin embargo, esta relación no es lineal, ya que el aumento del número de personas que visualizan un contenido implica una relación social entre esas personas, es decir, una afinidad en sus intereses, y por lo tanto una mayor correlación entre los intereses del grupo (que es el que consume) y los intereses individuales.

La fórmula para modelar este factor, en la que habría que ajustar h , i y j con valores reales del usuario o grupo de usuarios, viene dada por (20):

$$F_{co} = j \cdot \frac{e^{-h \cdot x}}{e^{-i \cdot x}} + 1 - j \quad (20)$$

Donde x es el número de usuarios que acompañan al usuario objeto de estudio en la visualización del contenido.

Es necesario introducir el parámetro j por la peculiaridad de afinidad entre grupos, que provoca que este factor de ponderación de compañía no pueda tener una asíntota en 0, ya que en ningún momento la compañía debe anular completamente la precisión.

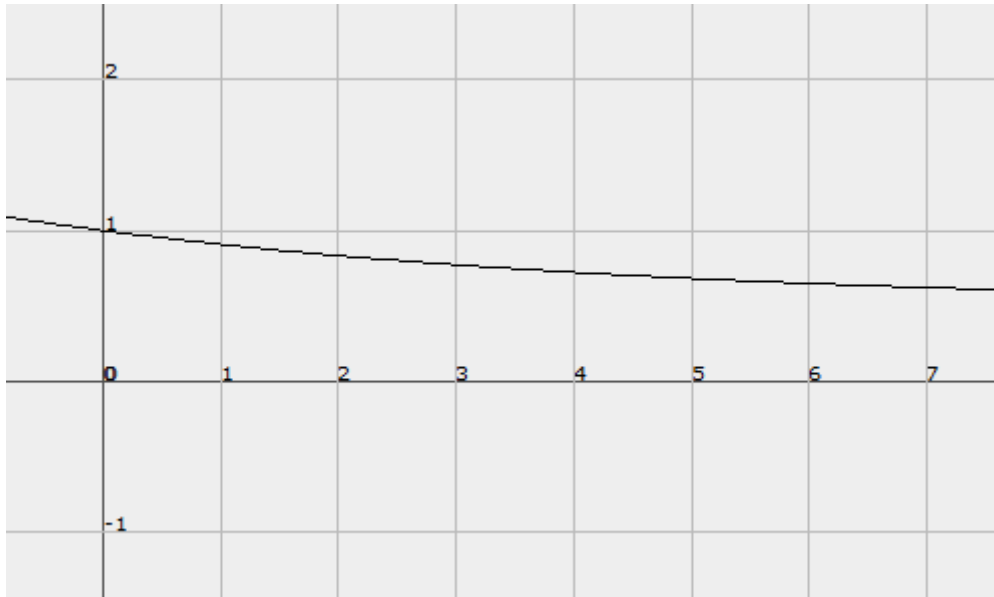


Fig. 18. Evolución del factor de compañía

La variable independiente corresponde al número de personas que visualizan un contenido simultáneamente, y la variable dependiente al factor de ponderación según compañía. Es necesario definir la pendiente y los puntos de inflexión, que dependen de los valores de h e i . Como ejemplo, se representa la gráfica de la Fig. 18 para los valores $h=0.4$; $i=0.2$; $j=0.5$.

Como se ha explicado en el modelo estadístico general, el cálculo del interés global lleva implícita la obtención de las correspondientes probabilidades de cada uno de los intereses posibles. Esta probabilidad da una idea de la certeza con que el interés inferido se acerca a la realidad. Sin embargo, para cada consumo individual es necesario tener en cuenta cómo influyen otros factores externos, como pueden ser, en este escenario *broadcast*, el factor de zapping o el factor de compañía ya explicados.

El impacto de estos dos factores, al estar cuantificados en una escala entre 0 y 1 y ser variables no correlacionadas, se puede conjugar mediante un simple producto, que se podría extender a otros factores externos que se pudieran considerar (21).

$$P'_n = \prod_{i=1}^N F_i \quad (21)$$

Este valor se utilizaría como parámetro de entrada para la etapa de cálculo del interés global, de la que saldrá un índice de validez, marcado por la probabilidad asociada a cada posible grado de interés. Este indicador marca la validez de un consumo individual dentro del cálculo del interés global, por lo que, para hallar la precisión general en cada momento hay que recurrir a un método incremental, ya que esta precisión, con cada nuevo dato aportado, nunca puede ser inferior. En el caso de que una medida de consumo no se considere válida debido a su alta imprecisión de entrada P'_n , la validez adoptará un valor mínimo que, al ser inferior al valor actual de la precisión de salida, no provocará cambio alguno.

De tal modo, la precisión definitiva P_n , partiendo de un valor inicial de probabilidad media (correspondiente al valor central de la distribución gaussiana que le corresponde), se irá incrementando paulatinamente con cada iteración, en el caso en que la validez sea

mayor que la precisión anterior, y este incremento será inversamente proporcional al número de iteraciones producidas hasta ese momento, es decir, al peso que tenga el consumo actual dentro de la cadena global de consumos. Esto se puede expresar mediante (22).

$$P_n = P_{n-1} + \left(V_n - \frac{P_{n-1}}{n} \right) \iff V_n > P_{n-1} \quad (22)$$

4.6 Aplicación del algoritmo a entornos multiatributo

Aplicación del modelo general y definición de atributos

Otra posibilidad de aplicación del modelo general se puede dar en entornos en los que cada contenido audiovisual está descrito mediante una serie de atributos prefijados. En este caso, el escenario de trabajo es el de acceso concurrente con recomendador únicamente en cliente, por lo que el consumo audiovisual del usuario debe mapearse en tiempo real en una serie de preferencias relacionadas con los atributos que describen los contenidos, y la recomendación final se genera del mismo modo en el terminal del usuario, sin intervenir el lado servidor en ningún momento.

La aplicación se ha realizado en este caso en un dominio de conocimiento concreto, pero sin necesidad de emplear herramientas semánticas. Concretamente, se trata de un recomendador de vídeos deportivos en el contexto de unos Juegos Olímpicos. Para ello, se caracteriza el gusto de los usuarios en función de cuatro atributos que corresponden con los metadatos que se envían junto a los contenidos: deporte, ronda, nacionalidad y deportistas. Estos atributos tienen distinta naturaleza, por lo que no se pueden modelar de forma homogénea.

1. Deporte. Representa la competición deportiva del vídeo, como fútbol masculino, baloncesto femenino o 100 metros lisos masculino. Su utilidad se debe a que cada usuario tiene sus deportes favoritos.
2. Ronda. Muestra la ronda de un partido o la fase de una competición, como final, semifinal, rondas previas... Su utilidad se debe a que algunos usuarios están especialmente interesados en las rondas finales.
3. Nacionalidad. Indica los países correspondientes a los deportistas que participan, o el país al que representan los equipos. Muchos usuarios están especialmente interesados en consumir vídeos deportivos en los que intervengan equipos de su propia nacionalidad.
4. Deportistas. Indica los nombres de los deportistas que participan en la prueba del vídeo. Es importante porque algunos usuarios están especialmente interesados en seguir a sus deportistas favoritos.

El sistema obtiene el valor de estos atributos a través de los metadatos que se transmiten asociados a cada vídeo a través de un archivo XML. Cada uno de los cuatro atributos definidos influye de distinta manera y en distinto grado sobre un usuario, por lo que el modelado del consumo audiovisual tiene como objetivo, en este caso, descubrir cuál es este grado de influencia sobre cada usuario. El sistema debe generar dos resultados de salida que servirán para realizar las recomendaciones finales:

- a) Attribute Relevance (AR). Expresa el grado de influencia que tiene un atributo sobre un usuario. Se mide en una escala de 0 a 1 y, por lo tanto, para cada usuario habrá cuatro valores, correspondientes a los cuatro atributos. La suma de estos cuatro valores debe ser 1. Este valor es equiparable, en el modelo general, al interés global.
- b) Attribute Value Certainty (AVC). Es necesario introducir este nuevo parámetro, ya que las preferencias de un usuario vienen dadas por el grado de influencia de un atributo (AR) ya mencionado y por los valores preferidos de ese atributo, estimados a través de la certeza de su valor. Los valores de los atributos son diferentes para cada usuario, y cada uno de ellos va acompañado de un valor de certeza (entero entre 1 y 5), obtenido de la cantidad de consumos asociados a un valor de atributo. Por ejemplo, para un usuario u_i , el atributo Deporte puede tener dos posibles valores (obtenidos de los consumos realizados), fútbol y baloncesto; si el primero de ellos es su deporte favorito, tendrá una certeza de 5; si al usuario le gusta el baloncesto, pero menos que el fútbol, tendrá un valor inferior, de 3 o 4. La medida del AVC se aplica a tres de los cuatro atributos, ya que uno de ellos, Ronda, puede ser completamente caracterizado con el valor de AR. Esto se debe a que, para un usuario al que importe la ronda de la competición que se disputa, siempre le interesarán más las rondas finales que las iniciales, y en caso de que no le importe no es necesario hacer ningún tipo de cuantificación. El atributo AVC se puede obtener directamente a través del consumo, como ya se ha comentado, pero también puede proceder de información explícita declarada por el usuario. Esta decisión queda en manos del diseñador del sistema y depende de sus requisitos y necesidades específicas.

Es importante distinguir ambos resultados, AR y AVC. Por ejemplo, se puede dar el caso de dos usuarios a quienes únicamente influya el atributo Deporte: este es su único condicionante a la hora de consumir un vídeo deportivo u otro; sin embargo, a uno de ellos sólo le gusta el fútbol y a otro sólo el baloncesto. El valor AR será el mismo para ambos usuarios, 5, mientras que el valor de AVC cambiará, siendo para el primer usuario 5 en fútbol y 1 en baloncesto y, para el segundo, 1 en fútbol y 5 en baloncesto.

La recomendación final de los contenidos dependerá del valor de estos dos parámetros, AR y AVC, por lo que resultan decisivos en la recomendación final.

En cuanto a la relación del modelo general con la adaptación a este entorno multiatributo, las similitudes son mucho mayores de lo que conceptualmente pueda parecer, y únicamente hay que añadir una pequeña etapa de adaptación de los consumos generales de un ítem a la significación de los atributos, como se explica en el siguiente apartado.

Arquitectura del sistema

La arquitectura del sistema de recomendación se divide en tres etapas, como muestra la Fig. 19. Las dos primeras son la adaptación del modelo de inferencia de interés desarrollado en esta tesis, mientras que la tercera etapa, de recomendación, resulta necesaria para comprobar el funcionamiento de las primeras. El proceso completo comienza con la captura de los datos de consumo ya especificada en la descripción del modelo (conjuntos de vectores U y V) y termina con la recomendación final.

La primera etapa parte de esos datos de consumo y obtiene una medida de interés local asociada a un contenido audiovisual (un vídeo deportivo, en este caso). A continuación se introduce en el sistema el modulo de adaptación, que transforma el interés local del dominio del vídeo al dominio de los atributos, dando una idea de la influencia de cada uno de los cuatro atributos y obteniendo también la medida del AVC. En este punto, el flujo del sistema se paraleliza en cuatro ramas, correspondientes a cada uno de los cuatro atributos del sistema. Aquí comienza la segunda etapa del sistema, que coincide con la obtención del interés global del modelo general. En este caso, este proceso, consistente en el Proceso Oculto de Markov (HMM) y en la etapa de inferencia bayesiana, se aplica en paralelo a los datos de consumo adaptados a cada uno de los cuatro atributos. Finalmente, la etapa de recomendación utiliza los valores de AR y AVC para filtrar y ordenar los ítems presentes en el corpus.

La arquitectura presenta un esquema similar para los dos casos de uso considerados, emisión en directo y emisión de programas almacenados. La única diferencia es que el escenario en directo añade un módulo final complementario que filtra los elementos recomendados dejando únicamente aquellos que estén en emisión en el momento en que el usuario solicita la recomendación. Los datos necesarios para este filtrado, fecha y hora de emisión de cada uno de los eventos, están incluidos junto con el resto de metadatos en los archivos XML previamente señalados.

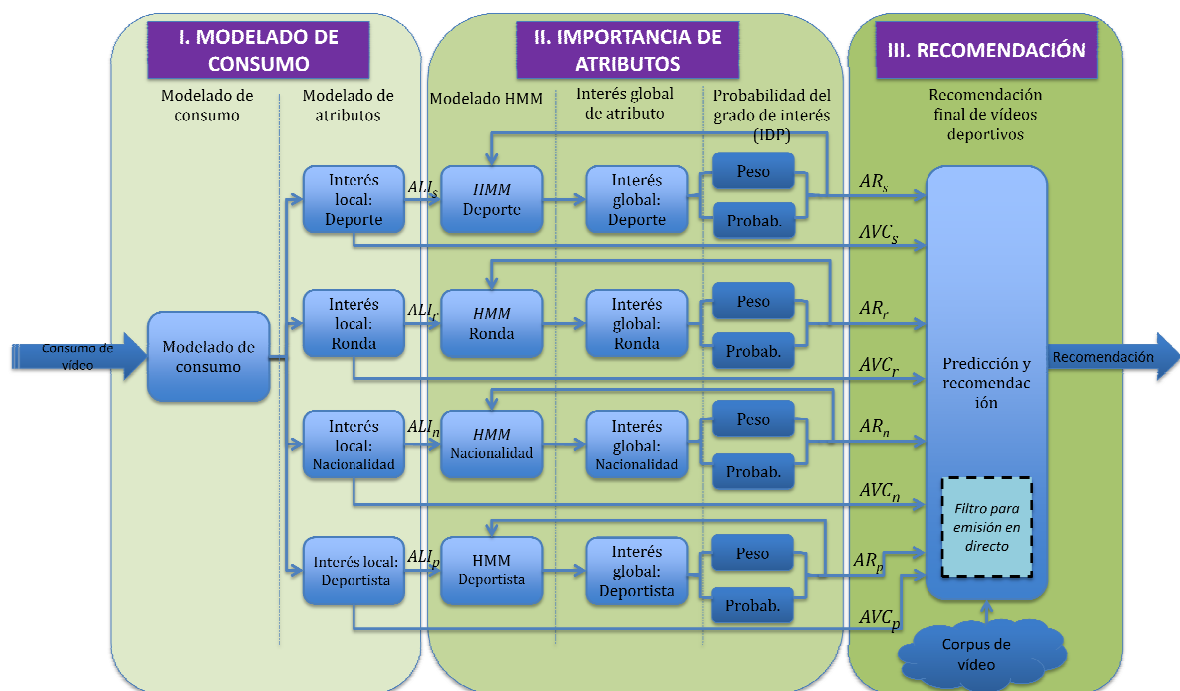


Fig. 19. Aplicación del modelo general a un entorno multiatributo deportivo

Modelado de interés local de atributos

El primer paso del sistema es la obtención del interés local del vídeo consumido a partir de los datos de consumo U y V . Para ello, se aplica el modelo general incorporando directamente un factor externo, denominado “*skimming and scanning*”, que equivale al factor de zapping de los entornos de transmisión de televisión convencional. Este parámetro representa la búsqueda por parte del usuario de fragmentos concretos dentro del vídeo, y viene dado por el número de *clicks*. Este parámetro traslada la degradación

del interés del usuario de forma análoga al del zapping y, una vez introducido en la fórmula, esta viene dada finalmente por la ecuación (23).

(23)

$$I_{LOC}=a+(1-a)\sum_{n=1}^N \frac{(v_n-u_n)}{\log_{10}(s)+1}$$

El valor final del interés local está comprendido entre a y 1. El parámetro a se ha introducido para fijar un valor mínimo de interés local para ítems consumidos, que en este caso debe ser diferente al de una transmisión de televisión convencional, puesto que aquí cualquier consumo responde a la voluntad del usuario, mientras que en una transmisión televisiva el usuario elige un canal, pero no un contenido específico. Además de este parámetro que establece un límite inferior de interés y de los valores de N y de los vectores v_n y u_n ya conocidos del modelo general, el parámetro s especifica el número de *clicks* del usuario dentro de la visualización de un vídeo, que pueden corresponder a saltos hacia adelante o hacia atrás en la barra de visualización, o a acciones de *fast-forwards* (avance rápido) o *rewinds* (retroceso).

La variación del interés local según el “*Skimming and scanning*”, dado por el número de *clicks*, se muestra en la gráfica de la Fig. 20 para diferentes tiempos de consumo de un contenido (expresado en % de la duración total del contenido). La ausencia de “*Skimming and scanning*” viene dada por el valor 1, que corresponde teóricamente al click inicial de *play* con el que se comienza cualquier visualización.

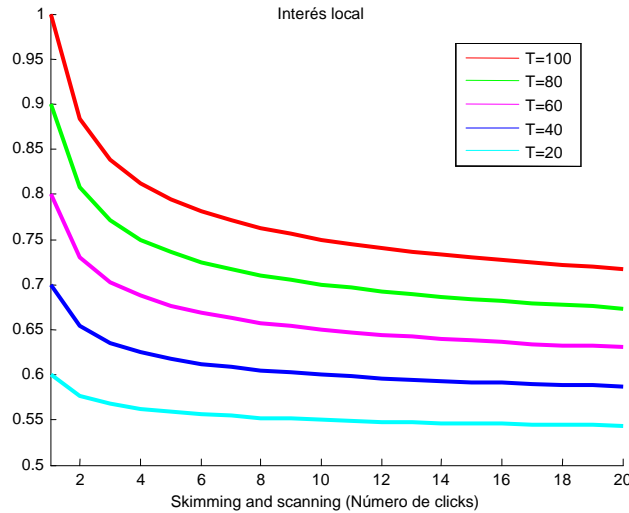


Fig. 20. Dependencia de la interrupción del consumo en el interés local

Una vez obtenido el valor de interés local asociado a un vídeo deportivo, es necesario introducir la etapa de adaptación al dominio de los atributos, que traduce el grado de interés en influencia de un atributo. Esta adaptación se realiza individualmente para cada uno de los cuatro atributos:

1. Deporte

La forma de saber la relevancia del atributo deporte es considerar aquellos que gustan al usuario con una probabilidad muy alta, ya se sepa mediante su información explícita o implícita. El valor de AVC se actualiza después de un consumo, y el interés local asociado al deporte viene dado por la ecuación (24), I_L es el interés local del vídeo y I_{LS} el interés local asociado al atributo Deporte.

$$I_{LS} = (-1)^a \cdot I_L \rightarrow a = \begin{cases} 1 \rightarrow \text{no está en perfil de usuario} \\ 2 \rightarrow \text{está en perfil de usuario} \end{cases} \quad (24)$$

2. Ronda

El modelado del interés para el atributo Ronda requiere (dado que no tiene sentido, como se ha comentado, utilizar valores AVC diferentes para cada usuario) de la definición de sus posibles valores (R). Se definen cinco valores dependiendo de la ronda (Preliminares o rondas clasificatorias: R=1; Segunda ronda: R=2; Cuartos de final: R=3; Semifinales: R=4; Final: R=5).

Si el atributo Ronda es importante para un usuario, esto implica que este prefiere consumir vídeos deportivos de las últimas rondas, es decir, con altos valores de R. Si, por el contrario, el atributo no le resulta influyente, la ronda no determinará el consumo del usuario y podrá presentar valores altos o bajos indistintamente.

El interés se modela mediante la ecuación (25), y se representa en la Fig. 21. La primera y última rondas son más representativas a la hora de estimar el interés del atributo porque la fiabilidad de la predicción aumenta (dan más información del atributo porque, por ejemplo, usuarios que ven muchas rondas iniciales lo hacen por motivos distintos a la ronda, mientras que usuarios que ven rondas finales pueden hacerlo por motivo de que las rondas sean finales y, por lo tanto, más importantes). Por lo tanto, la pendiente correspondiente a rondas extremas es más elevada.

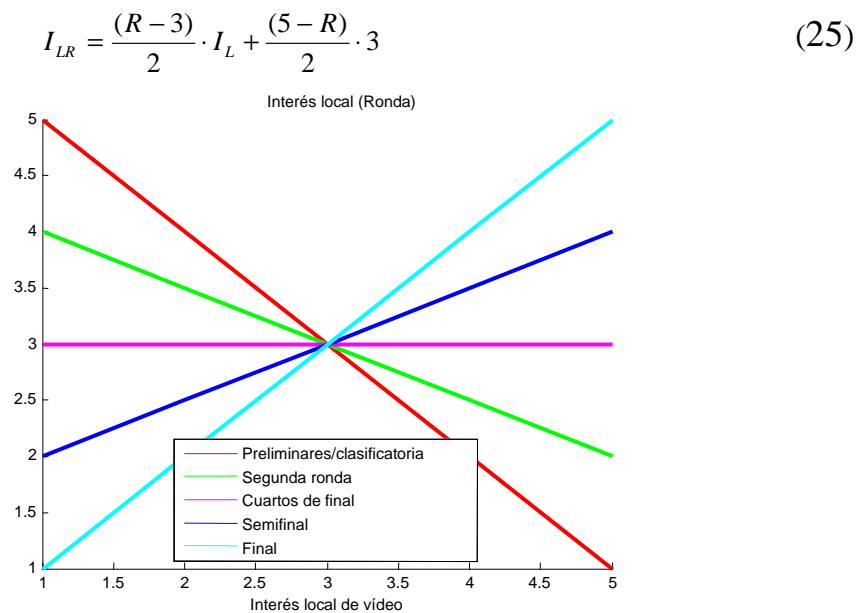


Fig. 21. Modelado del atributo "Ronda"

3. Nacionalidades

El modelado del atributo Nacionalidad es análogo al del atributo Deporte pero, a diferencia de este (que solo se aplica para deportes presentes en el perfil del usuario), siempre se aplica, dado que sólo se considera que la nacionalidad del usuario esté o no presente en el contenido del vídeo. Por lo tanto solo hay dos casos posibles, la presencia de algún deportista de la nacionalidad del usuario o de ninguno de ellos, lo que se expresa en la ecuación (26).

$$I_{LN} = (-1)^a \cdot I_L \rightarrow a = \begin{cases} 1 \rightarrow \text{otra nacionalidad} \\ 2 \rightarrow \text{nacionalidad propia} \end{cases} \quad (26)$$

4. Deportistas

Por último, el atributo Deportistas se modela de forma completamente análoga a Deporte, sustituyendo sólo el tipo de datos del perfil con los que se compara.

$$I_{LA} = (-1)^a \cdot I_L \rightarrow a = \begin{cases} 1 \rightarrow \text{no está en perfil de usuario} \\ 2 \rightarrow \text{está en perfil de usuario} \end{cases} \quad (27)$$

Finalmente, el modelado de interés local de los atributos genera dos salidas para cada uno de estos atributos: el *Attribute Value Certainty* (AVC) y el *Attribute Local Interest* (ALI). Ambos se cuantifican, como ocurría en el modelo general, mediante un cuantificador lineal, obteniendo una salida discreta. El parámetro ALI se utiliza como entrada del módulo de inferencia del interés global, ya explicado en el modelo general, que entregará como salida el valor de importancia del atributo para el usuario (AR).

Por otro lado, el parámetro AVC se utiliza directamente, junto con el valor de AR, en la etapa de recomendación, ya que indica los valores preferidos del usuario para cada atributo.

Etapas de recomendación

Finalmente, la tercera etapa del sistema, de recomendación, se implementa para poder generar resultados finales y realizar las correspondientes pruebas de funcionalidad, dado que el sistema de recomendación se basa, fundamentalmente, en el sistema de modelado de consumo ya expuesto.

El acceso por parte del usuario a la etapa de recomendación es independiente de las etapas previas, pero estas son necesarias para generar los datos que se utilizan para crear las recomendaciones.

En esta etapa de recomendación, para cada usuario el sistema genera un *rating* correspondiente a cada uno de los vídeos que todavía no ha consumido. Este *rating* (W_T) se genera mediante la suma de los pesos parciales de las predicciones de interés de los cuatro atributos que caracterizan cada vídeo, como se expresa en (28)

$$W_T = W_R + W_N + W_S + W_P \quad (28)$$

A continuación, se presentan las expresiones que caracterizan los pesos parciales de cada atributo, estableciendo la relación entre preferencia del usuario y metadato del contenido. Estas fórmulas se han obtenido parametrizando los datos de entrada para obtener fórmulas generales, de tal forma que se adapten a las diferentes escalas elegidas por los diseñadores. Los parámetros L_{ha} y L_{la} corresponden a los extremos superior e inferior utilizados para fijar la escala de la variable AR. L_{hw} y L_{lw} , a su vez, representan los límites de la escala de los pesos de los atributos de salida. Típicamente, estos límites son los mismos para los cuatro atributos, pero es posible modificarlos en ciertas aplicaciones para dar mayor importancia a unos u otros atributos. AR_a y AVC_a son los valores ya especificados de importancia y valor del correspondiente atributo a . El perfil de usuario, definido en (29), es un vector de conjuntos correspondiente al usuario p , que contiene información acerca de la nacionalidad del usuario o sus nacionalidades favoritas ($N_p \subset N$, donde N es el conjunto de todas las posibles nacionalidades), sus deportes favoritos ($S_p \subset S$, donde S es el conjunto de todos los deportes posibles), y deportistas favoritos ($A_p \subset A$, donde A es el conjunto de todos los posibles deportistas).

$$U_P = (N_p, S_p, A_p) \quad (29)$$

Ronda:

$$W_R = L_{lw} + (L_{hw} - L_{lw}) \cdot \left(\frac{R-1}{n-1} \right) \cdot \left(1 + \frac{AR_R - 1}{L_{ha} - L_{la}} \right) \rightarrow \forall R \in Z[1, n] \quad (30)$$

Donde R es el valor de la ronda para un evento deportivo dado y n es el número total de rondas consideradas en el diseño del sistema.

Nacionalidad:

$$W_N = \left(\frac{L_{hw} - L_{lw}}{L_{ha} - L_{la}} \right) \cdot AR_N \rightarrow \exists n_i \in N \mid \{n_i\} \subset N_p \quad (31)$$

$$W'_N = L_{lw} + \frac{1}{(P-1)} \cdot \left(\frac{L_{iw} - L_{hw}}{L_{ha} - L_{la}} \right) \cdot (AR_N - L_{la}) \rightarrow \forall N \not\subset U_p$$

Donde N es el conjunto de las nacionalidades de los deportistas participantes de los eventos deportivos, N_p es el conjunto de las nacionalidades preferidas (habitualmente, sólo la propia nacionalidad del usuario), y P es la penalización ($p \in Z[1, m]$) si no hay competidores del país del usuario.

Deporte:

$$W_S = L_{lw} + (L_{hw} - L_{lw}) \cdot \left(\frac{AVC_S - 1}{m-1} \right) \cdot \left(1 + \frac{AR_S - 1}{L_{ha} - L_{la}} \right) \rightarrow \forall AVC_S \in Z[1, m] \quad (32)$$

$$W'_S = 1 + \frac{1}{P} \cdot \left(\frac{L_{hw} - L_{lw}}{L_{ha} - L_{la}} \right) \cdot (AR_S - 1) \rightarrow \forall S \not\subset U_p$$

Donde m es el número de posibles valores de AVCs, U_p el perfil de usuario, S el deporte del evento, y P la penalización ($P \in \mathbb{Z}^{(1,m]}$) si el deporte no está en el perfil del usuario.

Deportistas:

El peso del atributo Deportistas es análogo al del atributo Deporte, siendo siempre $P=1$.

En la implementación del sistema, los valores AVC sólo se han empleado en los casos de Deporte y Deportistas, ya que para Ronda y Nacionalidad se establecen de forma implícita y genérica para todos los usuarios (en el caso de nacionalidad, el valor corresponde a la nacionalidad propia del usuario, obtenida como información explícita en el momento del registro)

4.7 Pruebas y validación

4.7.1 Pruebas en entornos categorizados

Implementación del modelo en un Set-Top box para la realización de pruebas

La realización de pruebas del modelo adaptado con usuarios reales implica la necesidad de disponer de un *software* de captura de los datos necesarios implantado en un dispositivo de consumo audiovisual. Para ello, se ha realizado una implementación del sistema en un receptor de TV digital (Set-Top box), empleando para ello las capacidades del middleware MHP (Multimedia Home Platform) [23] para el desarrollo de aplicaciones que puedan ser lanzadas por el usuario o se abran automáticamente.

La implantación del modelo en un sistema de recomendación empotrado en un dispositivo electrónico de consumo audiovisual se muestra en la Fig. 22.

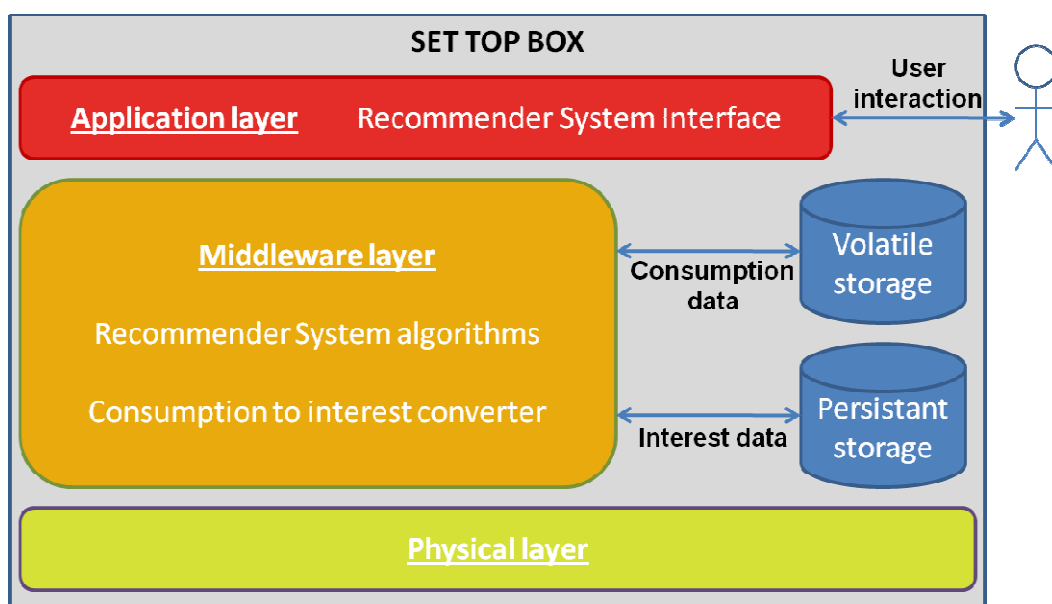


Fig. 22. Implementación en un Set-Top box

Sin embargo, para la realización de las pruebas, el módulo de conversión de consumo en interés no se ha ubicado en el dispositivo cliente, sino en el lado servidor, ya que esto permite que, en esta etapa desarrollo experimental, exista un mayor control y posibilidades de ajuste del sistema. El posterior desarrollo del sistema en una etapa comercial supondría la introducción del sistema completo en un dispositivo electrónico y no presentaría problemas, como se presenta para el caso de un Set-Top box en la Fig. 22.

El dispositivo dispone de una capa *middleware* que opera directamente sobre la capa física del receptor. La aplicación captura los datos de interacción del usuario como si fuera un medidor de audiencias convencional, y estos datos se almacenan en la memoria volátil del dispositivo. Cuando se almacenan los datos de consumo de un ítem local, se ejecuta el proceso de conversión del consumo en interés.

Dado que existe un conjunto de ítems globales definidos, el correspondiente interés del usuario por cada uno de ellos se almacena en la memoria persistente del receptor, dado que, a diferencia del interés local, este valor será el que sirva para la realimentación y actualización de los valores de interés del sistema. Para la ejecución del algoritmo es necesario almacenar sólo unos pocos datos en la memoria persistente: los valores de las matrices A y B y las medias y varianzas de las distribuciones de sus filas, el último valor de interés global y el último valor de precisión. Estos datos siempre están presentes en la memoria persistente del dispositivo y se actualizan con cada iteración del algoritmo (cada vez que se produce un consumo local). Finalmente, los datos de interés global del usuario se almacenan en la memoria persistente y sirven de entrada al motor de recomendación.

Hay dos casos de aplicación posibles dependiendo del motor de recomendación que se desee usar (y que las características del escenario permitan):

- a) Si se emplea un recomendador basado en contenido se generan las recomendaciones en el dispositivo, por lo que los valores de interés global se leen directamente de la memoria persistente.
- b) Si se emplea filtrado colaborativo o algún tipo de recomendación basado en técnicas sociales, los datos de interés global se envían al servidor empleando la conexión de red del dispositivo (Ethernet en el caso de los Set-Top-boxes con los que se ha trabajado). Se ejecutan los algoritmos de recomendación en el servidor y los resultados se envían de vuelta al dispositivo del usuario

Entorno de pruebas

El objetivo de las pruebas es comprobar la precisión del algoritmo de conversión de consumo audiovisual en interés, para lo que se procede a la comparación de los *ratings* inferidos por el sistema a partir del consumo audiovisual con las valoraciones manifestadas explícitamente por los usuarios que participan en las pruebas. A continuación se detalla el procedimiento seguido:

En primer lugar, es necesario tomar, de cada usuario incluido en la muestra, medidas exhaustivas de su consumo televisivo. Para ello, se implementa en los Set-Top-boxes que se entregan a los usuarios un *software* de captura de datos; en paralelo, se elige una serie de ítems globales de la parrilla de TV, cuya emisión coincida en el tiempo con el

período de pruebas en el que los usuarios dispondrán del Set-Top-box. A continuación, los datos de consumo almacenados se utilizan para inferir los *ratings* de cada usuario sobre cada ítem global. Cuando termina el período de prueba con los usuarios, se les pasa una encuesta pidiendo una valoración en una escala de 1 a 5 de los ítems globales seleccionados, de acuerdo con sus preferencias. Finalmente, se obtienen medidas de correlación entre los valores inferidos y los manifestados por los usuarios para validar el funcionamiento del sistema.

Captura de datos de consumo

El primer y más importante paso del procedimiento es la captura de datos reales de consumo de los usuarios, para luego aplicar los algoritmos directamente sobre ellos. Por esta razón, para el desarrollo de las pruebas, se partió del desarrollo de un sistema de medición de audiencias sobre MHP, y se modificó y adaptó adecuadamente según las necesidades del sistema. Los datos de captura, sin embargo, han servido íntegramente, aunque se haya tenido que modificar su formalización y la manera de expresarlos. El desarrollo de la implementación MHP [23] garantiza la flexibilidad de la solución que, por lo tanto, también podría desarrollarse en sistemas actuales más potentes, como HbbTV [25]. Resulta muy importante, de cara a una correcta captura de los datos, que la aplicación desarrollada cumpla todos los requisitos funcionales y de usabilidad necesarios, con el fin de que no se pierdan los datos y que el usuario no observe comportamientos anómalos sobre su visualización habitual del televisor. De cara al usuario, una vez se ha identificado en el sistema, el funcionamiento debe ser totalmente transparente. Por lo tanto, se han validado las funcionalidades y prestaciones del sistema haciendo uso de una plataforma [26] creada para facilitar el desarrollo y despliegue de aplicaciones interactivas en MHP, algo que se ha hecho con otro tipo de servicios, como se explica en [27].

El proceso de captura de datos comienza cuando el usuario enciende su Set-Top box. En ese momento, aparece una pantalla que muestra los nombres de los miembros de la casa y se pide al usuario su identificación, tal y como muestra la captura de pantalla de la Fig. 23. Para ello, lógicamente, es necesaria una sencilla configuración previa, en la que el instalador introduce los nombres de los habitantes de la casa en un sencillo archivo XML.

Una vez que el usuario o usuarios están consumiendo contenidos audiovisuales, el sistema almacena toda la información de consumo de forma estructurada en una serie de archivos XML. De esta forma, la memoria persistente del receptor almacena los valores de los conjuntos de vectores U y V descritos en (4), necesarios para la ejecución del sistema de inferencia del interés. Dado que el escenario de aplicación del sistema es el de difusión de televisión en directo y que el dispositivo empleado no es un PVR (*Personal Video Recorder*), no existen modos de visualización anormal, por lo que no es necesario tener valores para los conjuntos W y X definidos en (6).

Para la realización de las pruebas, se instalaron los receptores (con la aplicación de captura) durante una semana en 5 viviendas diferentes. En total, 18 personas, correspondientes a esas 5 viviendas, participaron en las pruebas, cuyos datos quedaron recogidos en 5 archivos XML [28].



Fig. 23. Identificación del usuario

Selección de ítems y encuesta

Para la realización de las pruebas, se han considerado dos tipos de ítems globales: “géneros de TV” y “programas de TV”. Para ello, se eligieron 10 géneros y 40 programas representativos de la parrilla de la televisión digital terrestre que se emitió en España durante la semana de realización de las pruebas. Los géneros fueron extraídos de las categorías definidas en TV-Anytime [23]. Por ejemplo, el ítem global “Deportes” se compuso de los ítems locales compuestos por las distintas retransmisiones de eventos deportivos. Para el otro tipo de ítems globales, los “programas de TV”, un ejemplo utilizado fue el ítem global “Telediario 2 TVE”, formado por los ítems locales de las emisiones diarias del telediario de la tarde de TVE. El resultado final del sistema es la inferencia de un rating por cada ítem global.

Una vez terminadas las pruebas, se pasó a los usuarios la encuesta que se muestra en el Anexo I de esta tesis, que se compone de dos partes, correspondientes a los dos tipos de programas tomados en consideración. En la primera parte de la encuesta, el usuario ordena los 10 géneros televisivos de acuerdo con sus preferencias, de mayor a menor. En la segunda parte, el usuario valora de 1 (menor interés) a 5 (mayor interés) los programas de TV que conoce. El diferente enfoque de las dos partes de la encuesta, decidido tras el estudio de las metodologías adecuadas [28], se debe a que los géneros son conceptos más generales, y se puede pedir al usuario una mayor precisión a la hora de compararlos, lo cual asegura una mayor cantidad de información recabada.

Una vez estaban recogidos los datos de consumo y los datos de las encuestas de los usuarios, se procedió a *parsear* automáticamente y extraer la información útil (U y V)

para introducirla en una base de datos MySQL. Por otra parte, la información de las encuestas fue introducida en otra tabla de la misma base de datos.

Finalmente, se implementó sobre un PC, trabajando sobre la base de datos implantada, un software con el modelo de conversión de consumo a interés propuesto en este capítulo. A partir de los datos de consumo, se generaron valores de interés local que, acumulados, dieron lugar a una serie de *ratings* de interés global, que se compararon con los resultados de las encuestas siguiendo las métricas definidas a continuación.

Pruebas y validación

La validación del sistema se realiza de dos maneras, mediante un análisis cualitativo (estudiando pormenorizadamente el resultado de cada uno de los usuarios), y mediante un análisis cuantitativo de la eficiencia de las predicciones (computando la correlación entre los *ratings* inferidos y las valoraciones explícitas de los usuarios).

b. Validación cualitativa

A continuación se presenta un ejemplo del funcionamiento del algoritmo para uno de los usuarios de la muestra. Para ello se ha elegido al usuario 9, ya que es uno de los usuarios que presentan un mayor número de eventos de consumo, lo que permite una comprensión más intuitiva de los resultados.

Ejecutando la simulación sobre los consumos del usuario 9, se obtiene un *rating* para cada ítem local y global. En primer lugar, se analiza la simulación del ítem global “Telediario 2 TVE”, que tiene lugar diariamente a las 21:00. La Tabla 7 muestra los resultados de interés global, así como su evolución a lo largo de la semana.

Evento	Interés local cuantificado	Interés global final	Validez	Precisión
Day 1	5	5	0.5703	0.5703
Day 2	3	4	0.4838	0.5703
Day 3	5	4	0.5710	0.5744
Day 4	1	3	0.2865	0.5744
Day 5	5	4	0.5233	0.5798
Day 6	4	4	0.5835	0.5878
Day 7	5	5	0.6771	0.5946

Tabla 7. Ejemplo de conversión de consumo a interés (Usuario 9): “Noticias TVE 1”

Como ya se ha explicado, en primer lugar se calcula el interés local del programa sobre su duración diaria. Este dato de salida sirve como entrada al algoritmo de cálculo del interés global, empleando además los factores de compañía y *zapping*. El factor de *zapping* se deduce fácilmente del número de medidas distintas que componen el consumo de un ítem local, mientras que el factor de compañía se obtiene de comparar las identidades de los usuarios que se identifican ante el Set-Top-box.

Como se puede comprobar en la Tabla 7, el primer interés local fija el valor inicial del interés global (ya que se utiliza para crear el vector inicial π del HMM). Los valores de interés local se mantienen altos, entre 3 y 5, la mayor parte de las veces. Por ello, el interés global se mantiene alto a pesar del bajo valor de interés local obtenido el día 4.

Esto significa que el sistema ha tratado este valor como un *outlier*, aunque en el momento en que se produce esto todavía no se sabe, ya que es necesario comprobar si ese valor supone un cambio de tendencia, hecho que finalmente no se produce. Por esta razón, el interés global baja ligeramente el día 4, al mismo tiempo que la medida de validez es muy baja, y la precisión, por lo tanto, no crece a pesar de contar con un dato más. Esto permite rectificar la tendencia rápidamente en las siguientes iteraciones.

El valor de precisión es una medida acumulativa: crece con los valores altos de validez, y se mantiene cuando la validez medida es baja (sea debido a un cambio brusco de tendencia o a unos factores externos, compañía y zapping, que añaden una excesiva incertidumbre). El valor de precisión nunca decrece, porque el sistema siempre tiene más datos, lo que significa más información o, en el peor de los casos, la misma información que en la iteración anterior. Por lo tanto, si los nuevos datos certifican la consolidación de una tendencia de consumo previamente inferida, la precisión crece. Si, por el contrario, el nuevo consumo no facilita nueva información relevante, el valor permanece o no crece significativamente. Un lento crecimiento del valor de la precisión también se puede deber a una alta influencia del Factor de Compañía o el Factor de Zapping.

Como se aprecia en el ejemplo, el sistema ofrece una forma general eficiente de interpretar los consumos audiovisuales e infiere a partir de los datos disponibles el grado de interés más probable que manifiesta un usuario sobre un ítem.

A lo largo de la semana de pruebas, se recogieron 119 consumos del usuario 9. Estos consumos corresponden con las 10 categorías de géneros televisivos previamente definidos. Dado que en la encuesta el usuario ordenaba los géneros de menor a mayor preferencia, se considera que la manera óptima de comprobar la eficiencia de la predicción es el uso del coeficiente de correlación de Spearman (como se explica en el siguiente apartado), cuya proximidad a 1 dará idea de la bondad de la aproximación. La Tabla 8 muestra los datos de interés del usuario 9 recogidos en la encuesta y las preferencias inferidas por el sistema, así como el valor total del índice de correlación.

	<i>Ranking de intereses explícitos de usuario 9</i>	<i>Interés inferido de usuario 9-Interés global</i>
1. Corazón	2	1
2. Series	8	4
3. Noticias	7	5
4. Deportes	6	3
5. Cine	9	5
6. Documental	10	5
7. Concursos	5	1
8. Reality shows	1	1
9. Infantil	4	3
10. Magazine	3	2
Coeficiente de correlación de Spearman	0.8789	

Tabla 8. Correlación entre preferencias implícitas y explícitas (Usuario 9)

c. Validación cuantitativa

1. Validación para géneros de televisión

Como ya se ha comentado para el usuario 9, los usuarios ordenan en la encuesta los 10 géneros televisivos de menor (1) a mayor preferencia (10). Estos datos se comparan con las puntuaciones de preferencia inferidas por el sistema, comprendidas entre 1 (menor preferencia) y 5 (mayor preferencia). La métrica elegida para comparar ambas clasificaciones es el coeficiente de correlación de Spearman, ya que este coeficiente permite cuantificar la dependencia estadística de dos variables que presentan una relación monótona. Se decide utilizar el coeficiente de Spearman en lugar del de Pearson porque la relación entre los datos no es lineal. Por ejemplo, dependiendo de los hábitos de consumo, ciertos géneros televisivos pueden corresponder a cuantificaciones absolutas más altas o más bajas respecto a una misma cuantificación relativa.

Los resultados obtenidos para los 18 usuarios se muestran con la línea continua roja de la Fig. 24, en la que el valor medio se representa mediante la línea roja discontinua, de valor $\rho_G=0.7817$.

2. Validación para programas de televisión

En la segunda parte de la encuesta, los usuarios puntúan los 40 programas televisivos definidos como ítems globales con una calificación entera comprendida entre 1 (menor interés) y 5 (mayor interés). En este caso, la encuesta expresa la misma característica de salida del módulo de inferencia de interés. Por lo tanto, hay una relación lineal entre ambos conjuntos de datos, lo que da lugar al uso del coeficiente de correlación de Pearson como métrica de comparación.

Los resultados obtenidos para los 18 usuarios se muestran con la línea continua azul de la Fig. 24, en la que el valor medio se representa mediante la línea azul discontinua, de valor $\rho_P=0.7027$.

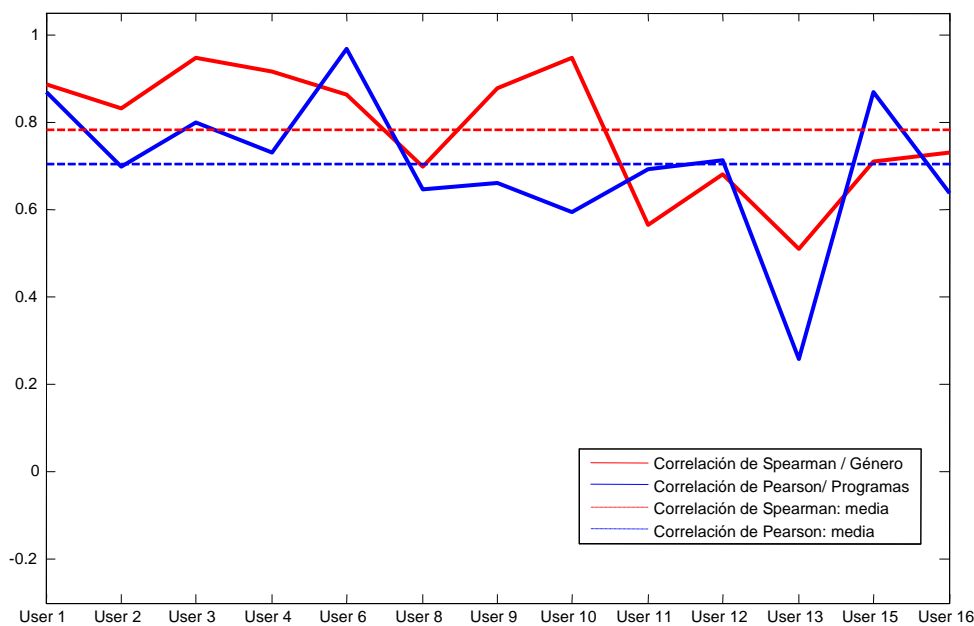


Fig. 24. Correlaciones de Spearman y Pearson para géneros y programas periódicos

Entre el conjunto de usuarios de la muestra, 5 de ellos fueron finalmente descartados por no haber realizado un número de consumos televisivos suficientemente representativo como para caracterizar sus preferencias. Concretamente, se descartaron aquellos de los que se recogieron menos de 10 consumos televisivos.

Los resultados concretos de correlación se muestran también en la Tabla 9. Los resultados de correlación se pueden comparar con el número total de eventos de consumo del usuario y con el número de consumos de “programas del corazón”. Se ha descubierto que los usuarios con más consumos en géneros “conflictivos” como el de los “programas del corazón” presentan una mayor inestabilidad en sus resultados de correlación.

Hogar	Usuario	Eficiencia género (Spearman)	Eficiencia programa (Pearson)	Núm. de eventos de consumo	Consumos del corazón
1	1	0.8857	0.8692	53	0 (0%)
1	2	0.8299	0.6992	14	0 (0%)
1	3	0.9476	0.7981	119	21 (18%)
1	4	0.9166	0.7303	98	10 (10%)
1	5	Descartado		1	0 (0%)
2	6	0.8620	0.9685	56	0 (0%)
2	7	Descartado		6	0 (0%)
2	8	0.6981	0.6455	19	0 (0%)
2	9	0.8789	0.6610	116	0 (0%)
3	10	0.9476	0.5944	64	3 (5%)
3	11	0.5636	0.6933	70	14 (20%)
4	12	0.6816	0.7115	68	23 (34%)
4	13	0.5085	0.2575	71	16 (23%)
4	14	Descartado		2	0 (0%)
5	15	0.7106	0.8700	122	7 (6 %)
5	16	0.7311	0.6366	35	1 (3%)
5	17	Descartado		6	0 (0%)
5	18	Descartado		9	0 (0%)
Media		0.7817	0.7027	69.6	-
Varianza		0.0208	0.0296	-	-

Tabla 9. Resultados de correlación (10 géneros / 40 programas)

En general, analizando los resultados, se comprueba que los géneros de televisión presentan mejores resultados que los programas de televisión, debido a que cada ítem global del tipo “géneros de TV” acumula en general un mayor número de datos que los ítems directamente relacionados con los “programas de TV”. A grandes rasgos, se aprecia un hecho evidente, que cuantos más consumos (más datos de entrada) la precisión del sistema aumenta. Además, se pueden establecer algunas conclusiones significativas muy interesantes:

- a) Se observa un comportamiento particular en ciertos hogares. Cuando hay varias personas viendo la TV al mismo tiempo, la eficiencia de los resultados difiere considerablemente de unos usuarios a otros. Esto se debe a que, en estos hogares, ciertas personas o grupos de personas dominantes eligen el consumo

televisivo de toda la familia. Por lo tanto, los resultados son menos eficientes para personas no dominantes. El “Factor de Compañía” explicado en apartados anteriores soluciona este problema, pero esto implica que el sistema requiera de más tiempo y consumos de los usuarios no dominantes para alcanzar una eficiencia de resultados similar al de las personas dominantes.

- b) Se puede deducir también que algunos valores bajos de eficiencia (como los correspondientes a los usuarios 11, 12 y 13) se deben a usuarios reticentes a manifestar sus preferencias reales por ciertos géneros conflictivos. Por ejemplo, se ha comprobado que los peores resultados se han obtenido en aquellos usuarios que consumen más programas del corazón (última columna de la Tabla 9), debido seguramente a que estos programas no poseen una buena reputación social y estos usuarios no admiten sus preferencias relativas a ellos. Se ha comprobado que la desviación más alta de los resultados de estos usuarios corresponden con altos consumos de programas del corazón y *reality shows* (que afirman que no les gustan), o bajos consumos de programas de buena reputación, como los documentales (que afirman que les gustan). Por lo tanto, se puede afirmar que algunos usuarios no son sinceros a la hora de expresar sus auténticas preferencias acerca de programas de reputación conflictiva. Con el objetivo de validar esta hipótesis, se ha realizado un nuevo experimento, y se ha eliminado del corpus de ítems globales tanto el género “programas del corazón” como todos aquellos ítems globales de programa relacionados con este género. Si los usuarios hubieran sido sinceros y este género no introdujera una distorsión significativa, los niveles de eficiencia deberían descender, ya que el sistema cuenta con menos datos para inferir las preferencias. Por lo tanto, se puede afirmar que los usuarios sinceros al manifestar su opinión respecto a los programas del corazón, presentan una disminución de la eficiencia al eliminar esta categoría (o, en todo caso, no presentan una mejora), mientras que los usuarios que no son sinceros, muestran una mejora muy significativa en cuanto a la eficiencia del sistema a la hora de predecir sus preferencias. Esta mejora se aprecia en las columnas correspondientes de la Tabla 10, y es muy significativa especialmente con los usuarios 11, 12 y 13. El cambio es tan significativo en estos usuarios que, calculando la media de las correlaciones, es resultado para todos los usuarios es superior al obtenido sin eliminar la categoría “programas del corazón”, a pesar de que la mayoría del resto de usuarios muestran un descenso de la eficiencia, como debería corresponder a una situación en la que se elimina información de entrada.

Se comprueba que, en ambos experimentos, los peores resultados corresponden al usuario 13. Analizando sus valoraciones en la encuesta y el resultado del proceso de inferencia de su interés, se comprueba que los dos problemas mencionados son los responsables. Durante la semana de pruebas, el usuario 13 ve la televisión acompañado la mayor parte del tiempo (lo cual provoca que se requieran más datos que para un usuario que consume televisión en solitario, como ya se ha comentado); además, declara que su género favorito son los documentales (que no ha consumido) y el que menos le gusta el de los programas del corazón (que ha consumido en cierto grado). Con más tiempo y más consumo televisivo, el usuario vería más programas en solitario y el sistema daría una mayor validez a sus eventos de consumo, que serían más representativos de sus gustos reales. Sin embargo, existen ciertos programas que en ningún momento admitiría que son de su preferencia. La suma de estos dos factores provoca que los resultados de este usuario sean los más deficientes.

Usuario	Eficiencia de género (Spearman)	Mejora de género (%)	Eficiencia programa (Pearson)	Mejora programa (%)	Consumo del corazón (%)
1	0.8365	-4.92	0.8629	-0.63	0
2	0.7979	-3.2	0.6876	-1.16	0
3	0.9084	-3.92	0.7981	0	18
4	0.8401	-7.65	0.6956	-3.47	10
5	Descartado				
6	0.8677	+0.57	0.9685	0	0
7	Descartado				
8	0.7182	+2.01	0.6455	0	0
9	0.8635	-1.54	0.6610	0	0
10	0.9820	+3.44	0.5944	0	5
11	0.7734	+20.98	0.7315	+3.82	20
12	0.7807	+9.91	0.7426	+3.11	34
13	0.6257	+11.72	0.3345	+7.70	23
14	Descartado				
15	0.6925	-1.81	0.8592	-1.08	6
16	0.6928	-3.83	0.6366	0	3
17	Descartado				
18	Descartado				
Media	0.7984	+1.67	0.7090	+0.63	-
Varianza	0.0098	52.89	0.026	12.16	-

Tabla 10. Resultados de correlación sin considerar programas del corazón (9 géneros / 36 programas)

Una conclusión importante a las pruebas realizadas con usuarios reales, derivada de esta observación relativa a los géneros y programas “conflictivos”, es que la información implícita puede resultar en ocasiones más fiable y representativa que la información explícita, aunque su uso dentro de sistemas reales, exclusivo o combinado, depende de muchos factores del escenario o la aplicación que van más allá de la eficiencia de los resultados, y que pueden estar relacionadas con la privacidad de la información o la confianza que se desee establecer con los usuarios. De todos modos, se ha comprobado la validez del modelo desarrollado a la hora de satisfacer una de las necesidades básicas de ciertos sistemas: el modelado automático del consumo audiovisual para descubrir los gustos reales de los usuarios. Estas preferencias son las que pueden servir como entrada para numerosos algoritmos de recomendación.

4.7.2 Pruebas en entornos multiatributo

Utilizando una implementación realizada del modelo adaptado definido en el apartado 4.6 para contenido deportivo, se han realizado varias simulaciones y pruebas que permiten validar de una forma cualitativa el funcionamiento del sistema. Unas pruebas completas con usuarios reales sólo serían posibles en el entorno real de unos Juegos Olímpicos, lo cual implicaría un despliegue comercial, con lo que se utilizan las simulaciones cualitativas de usuarios como referencia.

Por claridad expositiva, se consideran solo tres atributos en los test desarrollados: Ronda, Deporte y Nacionalidad. El comportamiento del atributo Deportistas es análogo al de Deporte, y requiere un alto número de consumos para que llegue a ser eficiente, ya que sería necesaria la existencia de muchos consumos en los que participara algún deportista determinado de forma recurrente.

Los límites de las escalas que se toman están comprendidos entre 1 y 5 para los parámetros AR y sus respectivas estimaciones de pesos, y el factor de penalización para conceptos no presentes en el perfil del usuario es $P=4$. Las ecuaciones resultantes de esta particularización para los atributos de Ronda, Nacionalidad y Deporte se especifican en (33), (34) y (35).

$$W_R = 1 + \left(1 + \frac{AR_R - 1}{4}\right) \cdot (R - 1) \rightarrow \forall R \in [1,5] \quad (33)$$

$$\left\{ \begin{array}{l} W_N = AR_N \rightarrow \text{nacionalidad propia} \\ W_N = 1 - \frac{1}{3} \cdot (AR_N - 1) \rightarrow \text{otras nacionalidades} \end{array} \right\} \quad (34)$$

$$\left\{ \begin{array}{l} W_s = AVC_s + \left(\frac{AR_s - 1}{4}\right) \cdot (AVC_s - 1) \rightarrow \text{está en perfil de usuario} \\ W_s = 5 - 4 \cdot AR_s \rightarrow \text{no está en perfil de usuario} \end{array} \right\} \rightarrow \forall AVC \in [1,5] \quad (35)$$

Para la realización de las pruebas, se generaron dos conjuntos de vídeos. El Corpus 1 consistía en 12 vídeos deportivos con diferentes características y sus correspondientes metadatos, y se utiliza como conjunto de entrenamiento para simular los consumos de los usuarios. En un escenario real, esto correspondería a los eventos pasados, que el usuario ha tenido la oportunidad de consumir en el pasado. Por su parte, el Corpus 2 (Tabla 11), que representa los eventos de próxima emisión, está formado por 30 vídeos con sus correspondientes metadatos, que se utilizan como ítems que se pueden recomendar a los usuarios.

Los sistemas comerciales trabajan habitualmente con un único corpus para entrenamiento y recomendación, pero la utilización de dos corpus, en este estado experimental del sistema, permite un mayor control, legibilidad y capacidad de comparación entre diferentes usuarios. Una condición previa de todo recomendador es que los vídeos ya consumidos por un usuario no se le puedan recomendar, algo que es implícito al uso de dos corpus. En el caso de utilizar un único corpus, la comparación de resultados entre usuarios sería imposible, ya que los vídeos que han sido consumidos por un usuario podrían ser recomendados a otros, con lo cual los resultados entre usuarios no serían comparables al no tener ambos los mismos posibles vídeos para recomendar. Por lo tanto, el uso de un solo corpus es adecuado para una fase de explotación, pero para la fase de pruebas y validación es necesario presentar dos conjuntos independientes.

Video	Deporte	Ronda	Nacion. 1	Nacion. 2
1	Football	Preliminaries	SUI	USA
2	Football	Preliminaries	CAN	NOR
3	Football	Preliminaries	BLR	FIN
4	Football	Preliminaries	GER	SWE
5	Football	Preliminaries	CZE	SVK
6	Football	Preliminaries	NOR	USA
7	Table tennis	Preliminaries	LAT	RUS
8	Basketball	Preliminaries	CAN	SUI
9	Basketball	Preliminaries	RUS	SVK
10	Basketball	Preliminaries	BLR	SWE
11	Basketball	Preliminaries	CZE	LAT
12	Water polo	Preliminaries	FIN	GER
13	Water polo	Preliminaries	NOR	SUI
14	Ice hockey	Preliminaries	LAT	SVK
15	Ice hockey	Preliminaries	BLR	GER
16	Ice hockey	Preliminaries	CZE	RUS
17	Ice hockey	Preliminaries	CAN	USA
18	Syncr. swimm.	Preliminaries	FIN	SWE
19	Badminton	Preliminaries	BLR	SUI
20	Beach volley	Preliminaries	CAN	GER
21	Tennis	Preliminaries	CZE	LAT
22	Tennis	Preliminaries	NOR	SVK
23	Judo	Quarter-final	SUI	USA
24	Football	Quarter-final	CAN	RUS
25	Table tennis	Quarter-final	CZE	FIN
26	Handball	Quarter-final	SVK	SWE
27	Archery	Semi-final	FIN	USA
28	Badminton	Semi-final	CAN	SVK
29	Handball	Bronze final	FIN	SVK
30	Football	Final	CAN	USA

Tabla 11. Corpus de recomendación

El último paso antes de comenzar las pruebas es la definición de los perfiles de usuario. Se crearon cuatro perfiles y se les asociaron características diferentes entre ellos, representativas de determinados patrones de comportamiento de grupos de usuarios y que permitieran comprobar diferentes funcionalidades del sistema de recomendación:

- El usuario 1 procede de Canadá, y le gusta ver competiciones en las que intervenga su país. Además, prefiere competiciones de las rondas finales.
- El usuario 2 también procede de Canadá, y le interesan especialmente algunos deportes concretos (sobre todo fútbol, pero también baloncesto). Este usuario también prefiere ver las rondas finales de las competiciones.
- El usuario 3 procede de Suiza, y le gusta especialmente ver las rondas finales de varias competiciones. También le gusta ver vídeos en los que haya representación de su país.

- El usuario 4 es español y su única preferencia es el baloncesto.

Según estas características, se crearon cuatro secuencias de consumo para cada uno de estos usuarios, sobre los vídeos deportivos del corpus 1, que se describen en la Tabla 12.

Usuar.	1	2	3	4
Nac.	CAN	CAN	SUI	ESP
Cons:	• Football	• Football	• Basketba	• Basketba
Deport	CAN vs. BRA	CAN vs. BRA	ll	ll
Nac.	Quarter final	Quarter final	LAT vs. CAN	LAT vs. CAN
Ronda	80%	100%	Final	Final
Loc	• Tennis	• Basketba	100%	100%
	USA vs. CAN	ll	• Badmint	• Basketba
	Semi-final	LAT vs. CAN	on	ll
	60%	Final	SUI vs. SWE	LAT vs. USA
	• Basketba	100%	Final	Preliminaries
	ll	• Football	100%	70%
	LAT vs. CAN	ARG vs. BLR	• Table	• Basketba
	Final	Quarter final	tenn	ll
	100%	80%	SUI vs. SWE	ARG vs. CAN
	• Ice	• Football	Semi-final	Semi-final
	hockey	SUI vs. BRA	100%	100%
	CAN vs. SUI	Quarter final	• Archery	• Basketba
	Semi-final	70%	USA vs. SWE	ll
	90%		Final	CZE vs. FIN
			70%	Preliminaries
				90%

Tabla 12. Simulación de consumos de usuarios

Estos consumos se realizaron desde diferentes perfiles de usuario en el sistema real, en el que se habían introducido previamente los vídeos del corpus 1. El sistema infirió la importancia de cada atributo para cada usuario (Tabla 13), y se pudo comprobar la correlación con la descripción cualitativa de los perfiles realizada previamente.

Atributo	Usuario			
	1	2	3	4
Ronda	3	3	5	2
Deporte	1	5	1	5
Nacionalidad	5	3	4	2

Tabla 13. Valores de AR (Importancia de atributo) inferidos por el sistema

Entonces, una vez que los perfiles de los usuarios quedaron almacenados, se sustituyó en el sistema el corpus 1 por el corpus 2, es decir, los eventos pasados se sustituyeron por los de próxima emisión, y cada uno de los usuarios obtuvo recomendaciones diferentes de los vídeos del corpus 2. Las primeras cinco recomendaciones de cada usuario se muestran en la Tabla 14.

Recomendaciones				
User	1	2	3	4
Pref 1	Video 30	Video 30	Video 30	Video 8
Pref 2	Video 28	Video 24	Video 23	Video 9
Pref 3	Video 24	Video 2	Video 27	Video 10
Pref 4	Video 20	Video 8	Video 28	Video 11
Pref 5	Video 17	Video 1	Video 1	Video 30

Tabla 14. Resultados de recomendación del sistema propio

Se puede comprobar, comparando los resultados de la Tabla 14 con la descripción del corpus de la Tabla 11 que las recomendaciones generadas son consistentes con los perfiles de usuario.

Estos resultados se extrajeron directamente de la interfaz del usuario. La Fig. 25 muestra una captura de pantalla de la interfaz, con la lista de recomendaciones obtenidas para el usuario 1 (a la derecha de la pantalla de visualización de los contenidos deportivos) mientras está consumiendo uno de los vídeos recomendados, correspondiente a Hockey sobre hielo.

La interfaz muestra el evento señalando los atributos tenidos en cuenta para la recomendación: el deporte, la ronda y los países participantes, con el objetivo (ya comentado en el capítulo 2 de esta tesis) de que el sistema sea transparente y el usuario no tenga que confiar ciegamente en lo que se le está recomendando, sino que compruebe las razones por las que esto sucede.



Fig. 25. Interfaz del sistema de recomendación: usuario 1

Con el objetivo de comparar el resultado del modelo propuesto en esta tesis con otros esquemas generales del estado del arte, se desarrolló un árbol de decisión que fue ejecutado sobre el mismo corpus y con los mismos perfiles de usuario. El árbol de decisión se muestra en la Fig. 26, y los resultados de la recomendación en la Tabla 15.

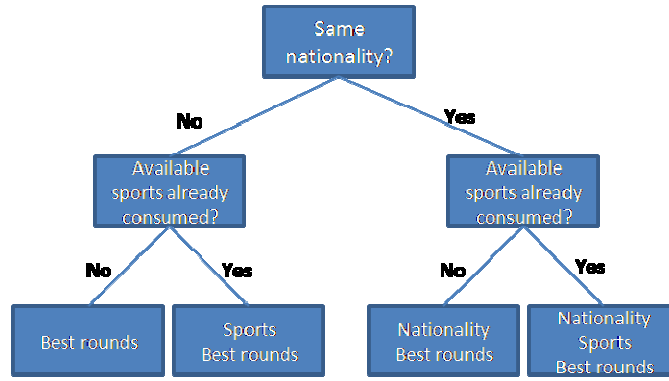


Fig. 26. Solución alternativa mediante árbol de decisión

Se puede comprobar que los resultados de la recomendación del árbol de decisión para los usuarios 2 y 4 son los mismos que los del modelo propuesto en la tesis. Sin embargo, los usuarios 1 y 3 reciben recomendaciones diferentes. La razón de esto es que el árbol de decisión genera buenos resultados para los usuarios que presentan un comportamiento típico, pero no es lo suficientemente flexible como para adaptarse a usuarios con gustos más anómalos, que precisan de un análisis más complejo. Estos usuarios necesitan un modelado flexible y dinámico de los atributos, como el que se desarrolla en el modelo propuesto.

Recomendaciones				
Usuario	1	2	3	4
Rec. 1	Video 30	Video 30	Video 8	Video 8
Rec. 2	Video 24	Video 24	Video 19	Video 9
Rec. 3	Video 17	Video 2	Video 23	Video 10
Rec. 4	Video 8	Video 8	Video 1	Video 11
Rec. 5	Video 2	Video 1	Video 13	Video 30

Tabla 15. Resultados de recomendación del árbol de decisión

Con el objetivo de obtener una medida cuantitativa de la precisión del recomendador, se genera una medida de correlación entre la importancia de un atributo para un usuario y la presencia de este atributo en la lista de recomendación final.

La importancia de un atributo en las listas de recomendación (RR_a) se obtiene considerando tanto la presencia del atributo como la posición en la lista de sus ítems relacionados.

Por ejemplo, la importancia del atributo Ronda en la lista de recomendación para el usuario 1 debería ser similar a su valor de AR ($AR=3$, como se puede comprobar en la Tabla 13). Para medir la importancia del atributo ronda en lista de recomendación se utilizó la ecuación (36), que dio para el usuario 1 un valor de $RR_r=4.27$.

$$RR_r = \left(\frac{\sum_{k=1}^{NR} NR \cdot pos_k \cdot pos_k}{\sum_{i=1}^{NR} pos_i \cdot NR} \right) \cdot \frac{A}{Max(RR)} \quad (36)$$

Donde NR es la longitud de la lista de recomendación ($NR=5$ en las pruebas) y pos_i la posición en la lista del ítem i . Por lo tanto, i representa todas las posibles posiciones en las listas de recomendación final (por ejemplo, si $NR=5$, i toma valores entre 1 y 5), y k

representa solo a los ítems relacionados con el atributo (por ejemplo, los deportes que están en el perfil del usuario o la propia nacionalidad del usuario). El término $A/Max(RR)$ es un parámetro de normalización ($A=5$ en este test, porque se utiliza una escala entre 1 y 5) y $Max(RR)$ es el máximo valor posible de la expresión dado el corpus (ya que hay que considerar que no siempre hay elementos disponibles que satisfagan las preferencias del usuario). En el ejemplo de ronda, el máximo valor de 5 solo se puede alcanzar si hay suficientes eventos de rondas finales en el corpus disponible.

Análogamente, las ecuaciones (37) y (38) estiman la importancia de los atributos deporte y nacionalidad en las listas de recomendación finales.

$$RR_S = \text{Max}_S \left(\sum_{k=1}^N \frac{(NR - \text{pos}_k + 1) \cdot L_S}{\sum_{i=1}^{NR} \text{pos}_i} \right) \quad (37)$$

$$RR_N = \left(\sum_{k=1}^N \frac{(NR - \text{pos}_k + 1) \cdot L_S}{\sum_{i=1}^{NR} \text{pos}_i} \right) \quad (38)$$

Utilizando los datos obtenidos de importancia de los atributos, que se pueden observar en la Tabla 16, se calculó la correlación con los valores de AR de cada usuario, tanto en el modelo propuesto como en el árbol de decisión implementado para la comparación. Los resultados se muestran también en la Fig. 27. Para el cálculo de la correlación se empleó el coeficiente de correlación de Pearson [30], y para los umbrales de correlación se eligió la definición de Johnston [31].

Usuario	Nacionalid.	Ronda	Deporte	Correlación (%)
1 (valor AR)	5	3	1	96.42
1 (Propio)	5	4.27	2.67	
1 (Árbol decisión)	5	3.47	3.33	94.37
2 (valor AR)	3	3	5	96.88
2 (Propio)	4.67	3.47	4.33	
2 (Árbol decisión)	4.67	3.47	4.33	
3 (valor AR)	4	5	1	92.43
3 (Propio)	1.67	4.68	2	
3 (Árbol decisión)	5	1.69	1.67	83.96
4 (valor AR)	2	2	5	99.78
4 (Propio)	-	1.53	4.67	
4 (Árbol decisión)	-	1.53	4.67	
Correlación media (Sistema propio)				96.38
Correlación media (Árbol de decisión)				93.75

Tabla 16. Valores obtenidos y correlaciones

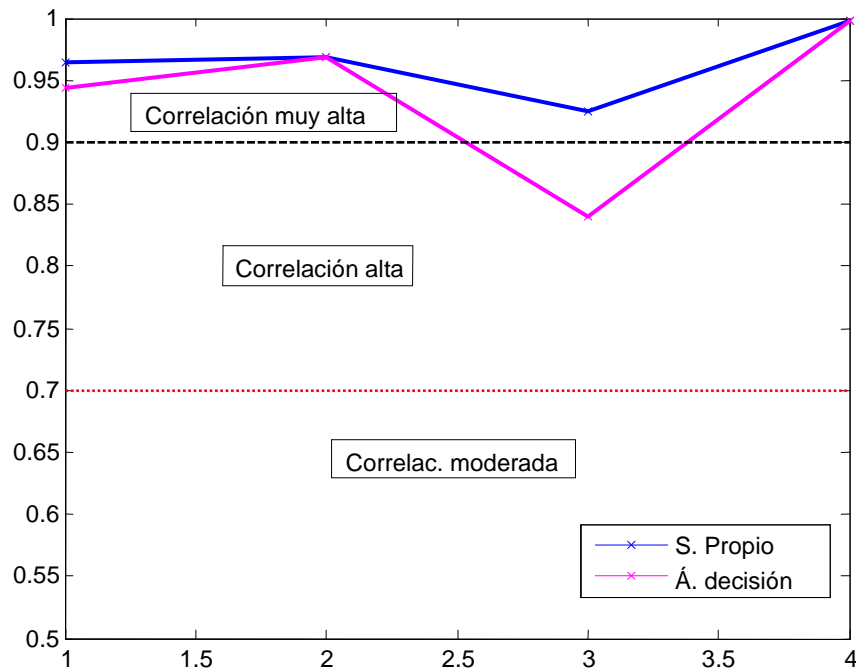


Fig. 27. Valores de correlación por usuario (umbrales de [31])

La correlación media de la propuesta de la tesis es mayor de 0.95, lo que significa que hay un alto grado de correlación entre los valores esperados y los obtenidos, y existe una correlación muy fuerte para los cuatro usuarios simulados. Por otro lado, como ya se ha señalado, el árbol de decisión funciona muy bien para usuarios más típicos, y algo peor con los usuarios con un comportamiento menos predecible a priori.

Finalmente, se muestra una tabla de comparación cualitativa entre la propuesta propia y otras posibles aproximaciones en la Tabla 17.

Los problemas de la propuesta definida en esta adaptación multiatributo del modelo general pueden aparecer en escenarios no controlados, ya que el sistema necesita un modelado específico de los atributos que caracterizan los contenidos. Un escenario controlado permite modelar previamente los atributos involucrados (en este caso de un escenario de los Juegos Olímpicos, los atributos de deporte, ronda, nacionalidad y deportistas en vídeos deportivos), pero un escenario no controlado implicaría la necesidad de incluir ontologías [35] o mecanismos de modelado semántico para generar dinámicamente el modelado de los atributos.

Otras propuestas	Ventajas de la propuesta de la tesis	Problemas de la adaptación al escenario de la aplicación de la propuesta alternativa
Martínez et al. [32]	<ul style="list-style-type: none"> • No necesita un número mínimo de usuarios. • Es más flexible, porque cada ítem está descrito utilizando varios atributos. • Tiene en cuenta comportamientos cambiantes del usuario, evaluando los cambios de tendencia en la etapa del HMM. 	<ul style="list-style-type: none"> • Es necesaria una implementación en el lado cliente. • Es necesario inferir preferencias de usuario. Esta aproximación pregunta al usuario de forma explícita.
Xiao Wu et al. [33]	<ul style="list-style-type: none"> • Tiene en cuenta todos los consumos de vídeo del usuario sin almacenar toda la información: el sistema se va actualizando en tiempo real. • Se toman en cuenta 4 características diferentes de los vídeos que son información descriptiva objetiva. La propuesta alternativa tiene en cuenta opiniones, links externos, información subjetiva. 	<ul style="list-style-type: none"> • Es necesario un modelo que tenga en cuenta preferencias de los usuarios, no solo similaridad de los ítems. • Es necesario un sistema que almacene preferencias de usuario que se puedan utilizar con contenido dinámico. Por ello, las preferencias de ítem se deben traducir a preferencias de atributo.
Shin et al. [34]	<ul style="list-style-type: none"> • Se infieren los gustos de los usuarios no solo del tiempo de consumo, sino también de la forma en que esta se produce. • Se asocian varios atributos a cada ítem, por lo que es más flexible y preciso. 	<ul style="list-style-type: none"> • Es necesario un modelo de consumo que tenga en cuenta escenarios televisivo y Web, con vídeos en directo y almacenados.

Tabla 17. Comparación con otras propuestas

4.8 Resultados obtenidos

En este capítulo se ha propuesto un modelo que permite extraer información de las preferencias del usuario a partir del análisis de su consumo audiovisual, con el objetivo de poder alimentar, a continuación, un sistema de recomendación de contenidos multimedia. Sobre el modelo propuesto, se han realizado pruebas con usuarios reales y usuarios simulados en diferentes escenarios, validando así la propuesta. A continuación se resumen los resultados específicos que se han alcanzado:

- 1) Se ha comprobado que el sistema propuesto funciona en diferentes escenarios, y con diferentes arquitecturas de recomendación. La ejecución del sistema puede tener lugar tanto en el lado del cliente como en el servidor. Además, se puede aplicar tanto a contenidos categorizados (como es el caso de la televisión) como a contenidos multiatributo (como ocurre en escenarios más específicos, como la prueba realizada sobre contenidos deportivos)
- 2) El módulo propuesto para la inferencia del interés local hace evolucionar el modelo clásico de tratamiento de la información implícita de consumo. El tiempo de consumo deja de ser el único indicador del interés de un usuario por un contenido, como hasta ahora proponía Hyoseop [14].
- 3) Se ha introducido el concepto de interés global, que permite aprovechar la redundancia entre distintos contenidos audiovisuales, ya venga dado éste por la pertenencia a una misma categoría o por la compartición de un determinado atributo. La propuesta de este concepto, además, se ha realizado con una gran flexibilidad, ya que dependiendo de las necesidades de cada aplicación es posible definir la relación entre ítems que caracteriza el interés global.
- 4) Se han realizado pruebas sobre los dos escenarios tratados, lo que ha permitido validar el sistema.
 - a. En el caso del escenario televisivo para contenidos categorizados, las pruebas se han efectuado con 18 usuarios reales repartidos en 5 hogares. Para ello, se midió exhaustiva y automáticamente el consumo audiovisual a través del dispositivo diseñado y, para validar, se compararon los valores de preferencia inferidos por el modelo con los manifestados explícitamente por los usuarios en una encuesta. Además, se ha demostrado que en ocasiones resultan más fiables las preferencias extraídas implícitamente del consumo del usuario que las manifestadas abiertamente, que pueden no ajustarse a la realidad por diferentes razones sociológicas y culturales.
 - b. En el caso de un escenario Web con contenidos definidos por una serie de atributos, se ha comprobado el funcionamiento del sistema mediante la ejecución de un conjunto de simulaciones sobre unos perfiles de usuario creados con ese objetivo.

5. Análisis y clasificación de contenidos audiovisuales para la recomendación

5.1 Introducción

Una vez que se ha conseguido modelar convenientemente la información del usuario, el desarrollo de un sistema de recomendación basado en contenido requiere también la catalogación y caracterización de aquellos ítems que se van a recomendar. En el caso de trabajo de esta tesis, el de los contenidos audiovisuales, es necesario, por un lado, crear un modelo de descriptores que sean útiles a la hora de definir los gustos de los usuarios y, por otro, definir unas metodologías de extracción y anotación automática de esos descriptores. Es decir, no es suficiente con definir unos descriptores teóricos que se compruebe que caracterizan las preferencias de los usuarios; esto debe ir acompañado de la posibilidad técnica de que esos descriptores sean extraídos de manera automática mediante técnicas que permitan su inclusión en un sistema de recomendación real.

Este capítulo aborda este doble objetivo, orientado tanto a la caracterización de las imágenes como de los vídeos, que son los dos tipos de contenidos que se consideran, en la presente tesis, representativos de los contenidos multimedia. Del mismo modo, se comprobará que algunas de las características de uno de los dos tipos de *media* pueden ser explotados en el otro.

Por lo tanto, tal y como muestra la Fig. 28 para el caso de los vídeos, el sistema de caracterización definido extraerá y modelará las características necesarias de un vídeo y las alojará en una base de datos, quedando a disposición, junto con las preferencias de usuario ya tratadas en el capítulo 4 de la tesis, de los algoritmos de recomendación (cuyo análisis se realiza en el capítulo 6).

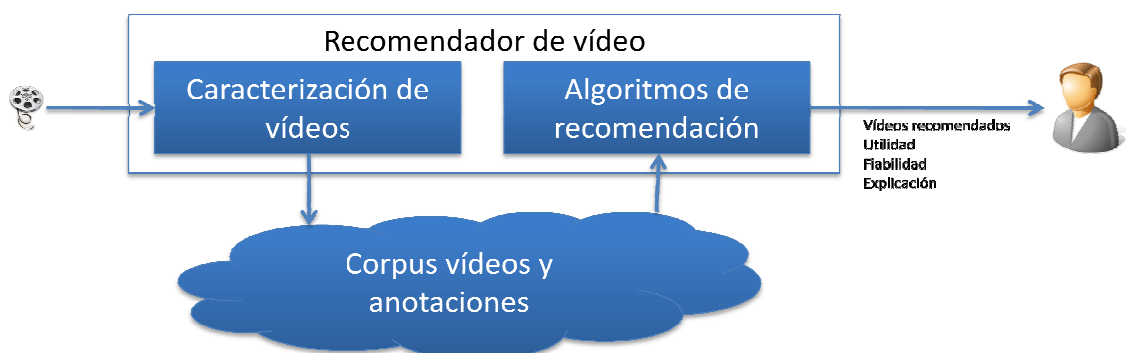


Fig. 28. Esquema general de recomendación de vídeos

Por lo tanto, desde el punto de vista de la arquitectura de los sistemas de recomendación, el proceso de caracterización del contenido se ejecuta de manera asíncrona al proceso de recomendación, ya que sólo necesita la disponibilidad del

propio contenido para comenzar sus operaciones, y la salida puede quedar almacenada hasta que los algoritmos de recomendación necesiten utilizarla.

Tipología de las características extraídas

Como queda reseñado en los capítulos introductorios de la tesis, uno de los principales objetivos es la creación de un modelo de recomendación basado en contenido que no esté supeditado al conocimiento semántico de dominios de conocimiento específicos. Desde este punto de partida, la definición de descriptores de imágenes y vídeo capaces de caracterizar las preferencias de los usuarios no puede hacer referencia a los temas tratados ni a los argumentos, contenidos o conceptos semánticos. Esta restricción deja una única vía de trabajo, que es la caracterización formal y estética de los contenidos audiovisuales.

El gusto audiovisual de los usuarios se define, como explicó el filósofo alemán Adorno en su Teoría estética [38], a partir de dos sistemas de coordenadas independientes: uno de ellos, su interés por los temas tratados o por los conceptos relacionados con el contenido audiovisual; el otro, la atracción o repulsión estética que les producen las formas en las que se basa un contenido audiovisual. Esto responde a la habitual clasificación de obras artísticas como resultado de la fusión de fondo y forma. Por lo tanto, el objetivo del presente capítulo es la caracterización formal y estética de los contenidos a partir de atributos de bajo nivel que se puedan extraer y modelar automáticamente. El especial interés de esta caracterización está en que las técnicas de recomendación vigentes basadas en contenido emplean sólo conocimiento semántico, y no hacen uso del que los criterios formales y estéticos pueden aportar.

5.2 Estado del arte

Caracterización de imágenes y de vídeos

Como base teórica para la categorización formal de las imágenes, se parte de los principios de estética del ya mencionado Adorno [38], de los enfoques psicológicos y cognitivos del estadounidense Gombrich [39], y de la actualización a los conceptos de estética contemporánea del francés Jacques Aumont. Las propuestas de los dos primeros autores se consideran especialmente para la caracterización de las imágenes. De Adorno se extrae la importancia en la estética de las imágenes de atributos como la textura, la luminosidad, el color y las proporciones y geometría (si bien, finalmente, no ha sido posible extraer automáticamente descriptores relacionados con este último punto, como se explica en el siguiente apartado). Por su parte, de las teorías de Gombrich se han obtenido ideas que han permitido definir descriptores específicos. Por ejemplo, la idea de la respuesta emocional del espectador ante los distintos colores, provocando emociones diferentes dependiendo de la saturación o de la distribución y homogeneidad de estos colores, se ha traducido en el modelo de descriptores de esta tesis en atributos específicos que cuantifican estas características.

A partir de estas ideas estéticas sobre la imagen, se han estudiado las aplicaciones más concretas de la materialidad de la imagen fija y de la imagen en movimiento llevadas a cabo por el propio Aumont en sus estudios sobre imagen [40] y cine [41], estos últimos complementados por las aportaciones de Alain Bergala, Michel Marie y Marc Vernet, y las contribuciones más centradas en las posibilidades tecnológicas de computación de la

percepción estética de Itten [42] y Davis [43]. El caso de Aumont es paradigmático y, como se explica en el apartado 5.4, se ha tomado como base para caracterizar estéticamente el vídeo. Por su parte, los estudios de Itten y Davis han servido para confirmar las teorías de Gombrich sobre el efecto subjetivo del color, ya aplicadas a características procesadas automáticamente.

Del mismo modo, también se ha estudiado la obra de otros pensadores de la imagen y el cine que aportan una visión más heterogénea, como los teóricos estadounidenses Noël Burch [44] y David Bordwell [45], que relacionan la narrativa audiovisual con ciertas métricas sobre la duración de planos y escenas en una obra fílmica, o los españoles Domonec Font [46] y Santos Zunzunegui [47], que encuentran patrones comunes relacionados con los movimientos de cámara y con el uso de determinados signos fílmicos (como transiciones, fundidos o cortes) en cinematografías y movimientos que comparten rasgos estilísticos.

Además del estudio teórico, se han analizado algunas herramientas que se utilizan ya en círculos académicos para los estudios de análisis fílmico. Concretamente, se ha encontrado la herramienta Cinemetrics [48], que permite obtener relevantes características de una película orientadas al análisis fílmico, lo cual guarda una relación muy estrecha con la caracterización estética del audiovisual y, por lo tanto, con la recomendación de contenidos multimedia. Cinemetrics es una herramienta manual que permite que un usuario, durante el visionado de una película, marque en una interfaz los instantes en los que se producen cambios de plano. En su versión avanzada (Fig. 29), además, permite indicar qué tipo de plano (plano general, plano americano, plano medio, primer plano...) es el que comienza. Finalmente, el programa obtiene una serie de estadísticos, que son los descriptores que el usuario que ha analizado la película puede compartir en la Web con la comunidad.

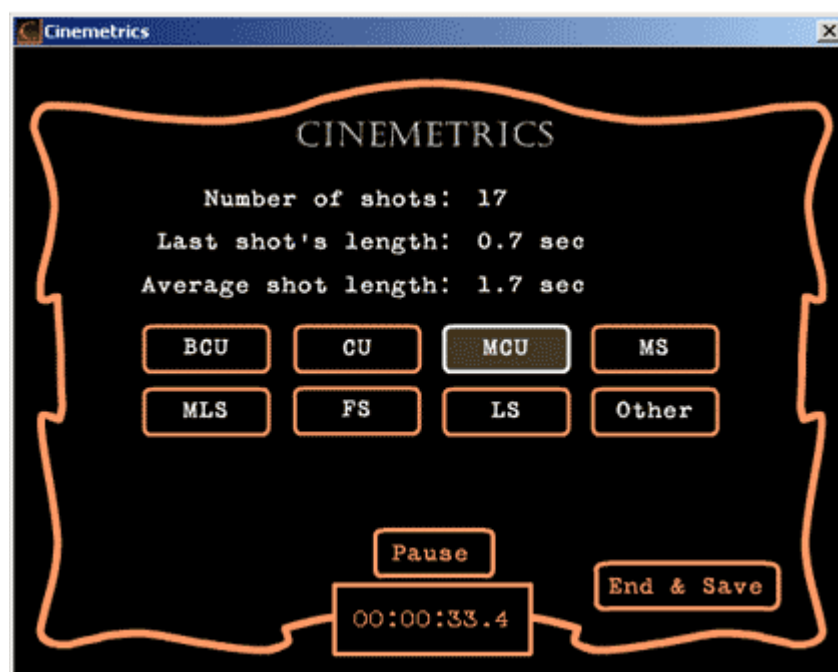


Fig. 29. Interfaz avanzada de Cinemetrics

El proceso de obtención de descriptores de Cinemetrics es tedioso, ya que requiere de la colaboración total de un visualizador, pero su éxito se debe al crecimiento de Internet, y la posibilidad de incorporar técnicas de *crowdsourcing*, consistentes en este caso en que

interesados de todo el mundo analizan películas y comparten los resultados en el portal Web de la herramienta. Importantes estudiosos y analistas fílmicos como David Bordwell o Warren Buckland hacen uso de esta herramienta, y la citan como importante fuente de información.

El objetivo de este capítulo de la tesis, además del esencial de crear un modelo de descriptores para la recomendación, consiste en la creación de una herramienta completamente automática, que permita analizar a gran velocidad un vídeo, resulte fiable (ya que no depende del factor humano) y que sea capaz de extraer descriptores que no son medibles a simple vista pero que influyen en la estética y en la percepción. Sin embargo, resulta interesante como punto de partida analizar cuáles son los descriptores empleados en Cinematics:

- ASL: Average Shot Length, duración media de planos, medida en segundos.
- MSL: Median Shot Length, duración mediana de planos, medida en segundos.
- MSL/ASL: Ratio entre ASL y MSL
- LEN: Length of the film, duración de la película, en minutos y segundos.
- NoS: Number of Shots, número de planos
- MAX: Plano más largo, en segundos
- MIN: Plano más corto, en segundos
- Range: Diferencia entre el plano más largo y el más corto
- StDev: Standard Deviation, desviación típica de la duración de los planos.
- CV: Coefficient of Variation, ratio StDev/ASL

Estos estadísticos se obtienen de cada película en general pero, además, para la aplicación avanzada, se aplican a cada uno de los tipos de plano que se han recogido.

Estándares de metadatos

Además de los estudios de análisis fílmico, también existen estándares técnicos que describen la adquisición y estructuración de metadatos para la descripción de contenidos audiovisuales. El más importante y potente de estos estándares (por su generalidad y porque hace uso de otros estándares anteriores, como Dublin Core [49]), es MPEG-7 [50].

El estándar MPEG-7, *Multimedia Content Description Interface*, está definido para recoger diferentes tipos de metadatos, tanto palabras clave como metadatos semánticos o metadatos estructurales. Este último tipo de metadatos es el que se aborda en esta tesis, por lo que se consideran las orientaciones de MPEG-7 para su extracción y se establece una correspondencia con su jerarquía que permita que el sistema sea interoperable con otros sistemas. Por este motivo, la parte más relevante del estándar, y que se analiza en profundidad, es la parte 3 (ISO/IEC 15938-3 [51]), que corresponde con las estructuras básicas y descriptores que cubren la información de tipo visual, tanto de imágenes como de vídeo. Como se comprobará en el capítulo, también se consideran, de forma más superficial, los descriptores de audio, recogidos en la parte 4 del estándar MPEG-7 (ISO/IEC 15938-4 [52]).

5.3 Modelo de extracción de características y anotación de imágenes para la recomendación

El objetivo del recomendador es el de proporcionar recomendaciones basadas en las características propias de las imágenes sin tener en cuenta información semántica de las mismas. Para ello es necesario realizar una caracterización de la manera general de definir y describir esas imágenes, en los términos deseados y de manera que se cree información valiosa a la hora de aplicar posteriormente los algoritmos de recomendación. Esta caracterización ha quedado plasmada en [36]. Por lo tanto, en primer lugar se ha realizado una definición y selección de los descriptores que se utilizan para caracterizar las imágenes de una manera general (independiente del gusto de cada usuario).

En el proceso seguido, con el fin de partir de materiales estandarizados, se ha comenzado realizando un estudio de los descriptores definidos en el estándar MPEG-7 [50] cuya información pudiera ser de especial interés en el proceso. Sin embargo, se ha comprobado que estos descriptores necesitan ser reformulados para poder extraer información directamente aprovechable. Del mismo modo, se han desarrollado descriptores nuevos que vienen a complementar la labor de los anteriores.

La Fig. 30 recoge todos los descriptores que finalmente se han empleado en el sistema de recomendación, dentro de lo que se ha denominado dominio D, de carácter general para todos los usuarios. Además, se presenta como ejemplo el proceso de extracción de uno de los descriptores, correspondiente a la homogeneidad de líneas.

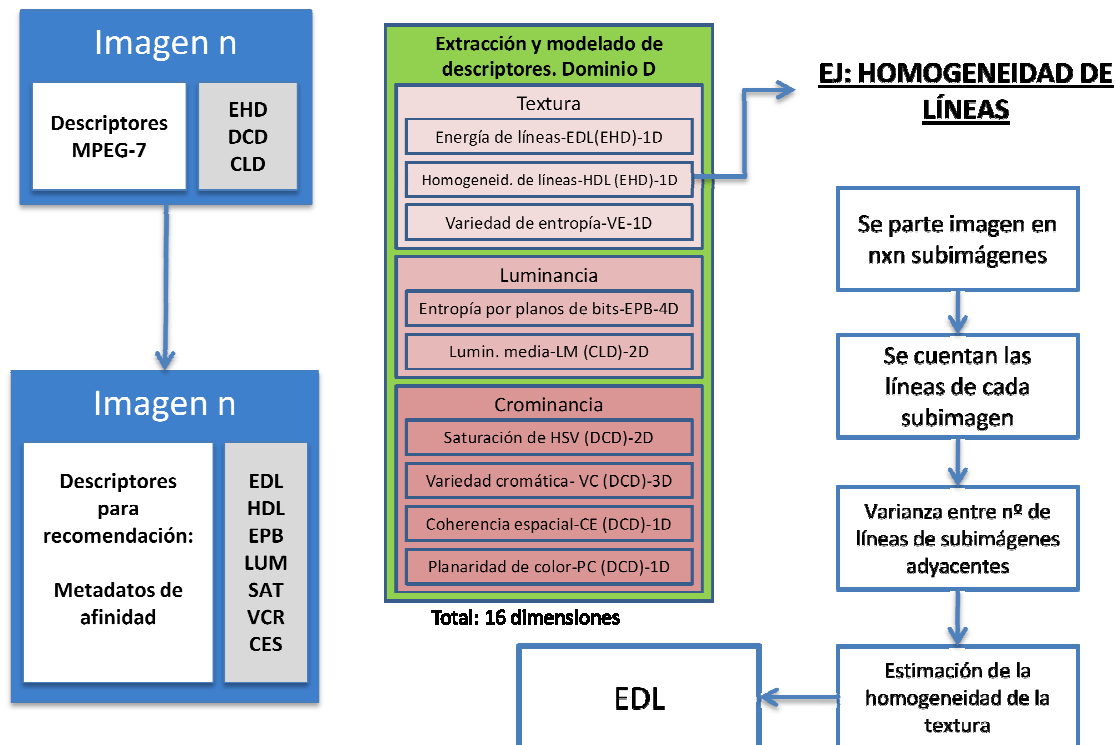


Fig. 30. Descriptores seleccionados para la clasificación de las imágenes y ejemplo de HDL

Como se aprecia en la Fig. 30, se han definido finalmente nueve descriptores, algunos de los cuales constan de varias dimensiones, por lo que al final cada imagen resulta representada por un vector de 16 dimensiones:

$$v = (E_1, H_1, VE, L_c, L_y, e_1, e_2, e_3, e_4, N, VC_{intra}, VC_{inter}, SC, S_m, S_y, PC) \quad (39)$$

Algunos de estos descriptores han sido obtenidos a partir de los tres descriptores de MPEG-7 que han resultado útiles, *EHD* (*Edge Histogram Descriptor*), *DCD* (*Dominant Color Descriptor*) y *CLD* (*Color Layout Descriptor*); otros, sin embargo, son directamente extraídos de las imágenes.

El siguiente apartado explicará brevemente las características de estos tres descriptores de MPEG-7 que han servido de base, mientras que el posterior apartado del capítulo se centrará en la explicación de cada uno de los nueve descriptores empleados en la fase de clasificación.

Descriptores base MPEG-7

EDGE HISTOGRAM DESCRIPTOR (EHD)

Se trata de un descriptor de texturas encargado de representar la distribución de bordes en la imagen a partir de su división en 4x4 subimágenes y la obtención de los histogramas correspondientes al reparto de bordes locales en cada una de ellas. Para generar dicho histograma los bordes son clasificados de acuerdo a su dirección: horizontal, vertical, diagonal 45°, diagonal 135° y sin dirección, obteniéndose con ello un total de 5x16=80 intervalos de histograma para cada imagen.

DOMINANT COLOR DESCRIPTOR (DCD)

Este descriptor de color se encarga de proporcionar una descripción compacta de los colores representativos en cada imagen, aportando:

- El número total de colores dominantes con un máximo de 8 e independiente para cada imagen.
- Su coherencia espacial, que representa la homogeneidad espacial de cada uno de los colores detectados.
- Su porcentaje o fracción de píxeles en la imagen correspondiente al color dado.
- Su varianza (opcional), encargada de indicar la fluctuación de los valores del color dentro de un cluster centrado en el color representativo detectado.
- Su valor en RGB.

El descriptor se proporciona en forma binaria, cumpliendo la siguiente sintaxis:

- Número de colores: 3 bits que indican el número total de colores dominantes detectado.
- Coherencia espacial: 5 bits.
- Porcentaje: 5 bits que indican el porcentaje normalizado asociado a cada color dominante.
- Varianza de color: 1bit.
- Valor: cada color dominante según su RGB.

En cuanto a su extracción, y tal y como se indica ampliamente en [53], se parte de un cluster único con un color representativo como centroide para cada imagen. A

continuación se realiza una secuencia de nuevos cálculos de centroides hasta alcanzar el criterio de parada elegido (mínima distorsión o máximo número de iteraciones). A partir de este punto, los clusters con mayor distorsión son divididos mediante la utilización de vectores de perturbación hasta que dicha distorsión descienda por debajo de un umbral o, en su defecto, hasta alcanzar el mayor número de clusters definidos.

En lo referente a la coherencia espacial, este parámetro se obtiene a partir del análisis de conexión de componentes, es decir, mediante la identificación de grupos de píxeles con el mismo color dominante que se hallen espacialmente conectados, haciendo uso para ello de una conectividad a cuatro (los cuatro vecinos más próximos a cada píxel).

COLOR LAYOUT DESCRIPTOR (CLD)

Este descriptor está diseñado para representar de forma eficiente la distribución espacial de colores de la imagen tratada. Se obtiene a partir de la aplicación de la transformada discreta del Coseno (DCT) a los colores de la imagen en el espacio YCrCb, previa división de ésta en 64 bloques. El proceso completo consta de los siguientes pasos:

- División de la imagen
- Detección del color representativo
- Aplicación de la DCT
- Cuantificación no lineal de los coeficientes escaneados en forma de zigzag.

Descriptores finales del clasificador para el sistema de recomendación

Según la naturaleza de los nuevos descriptores generados, estos se han clasificado en tres grandes bloques: textura, luminancia y crominancia.

1. Textura: Energía de líneas (EDL)

A partir del descriptor EHD se ha diseñado este parámetro que pretende medir la densidad de líneas en la imagen a partir de su energía con el objetivo de poder diferenciar aquellas imágenes con una numerosa cantidad de líneas de otras en las que no haya apenas transiciones.

Este parámetro está compuesto por un único valor que aporta la información obtenida, y que se halla de la siguiente forma:

$$E_{lTot}^2 = E_{45^\circ}^2 + E_{90^\circ}^2 + E_{135^\circ}^2 + E_{180^\circ}^2 + E_{Other}^2 \quad (40)$$

2. Textura: Homogeneidad de líneas (HDL)

Desarrollado también a partir del descriptor de MPEG 7 EHD, este descriptor se encarga de estudiar la continuidad de las líneas en la imagen. Cuenta también con una única dimensión H_1 , cuyo valor se obtiene dividiendo la imagen original en 16 subimágenes. Entonces, se halla la varianza de la distribución de líneas entre las 4x4 subimágenes vecinas, como muestra la siguiente expresión:

$$H_1 = V_1 + V_2 + V_3 + V_4 + V_5 \quad (41)$$

Donde:

V_1 = varianza entre las subimágenes (1,1),(1,2),(2,1),(2,2)
 V_2 = varianza entre las subimágenes (1,3),(1,4),(2,3),(2,4)
 V_3 = varianza entre las subimágenes (3,1),(3,2),(4,1),(4,2)
 V_4 = varianza entre las subimágenes (3,3),(3,4),(4,3),(4,4)
 V_5 = varianza de V_1, V_2, V_3, V_4 .

Y teniendo en cuenta la distribución de líneas a lo largo de cada dirección.

3. *Textura : Variedad de entropía (VE)*

Este parámetro representa la varianza de la entropía a lo largo de toda la imagen, y se calcula dividiendo la imagen en un conjunto de 5x5 subimágenes y obteniendo el valor de la entropía para cada uno de ellos. El parámetro final es la varianza entre esos 25 valores de entropía localizada.

4. *Luminancia: Entropía por planos de bits (EPB)*

Con este descriptor se pretende hacer un estudio de la distribución de entropías según planos de bits, de acuerdo a lo explicado en [55]. El procedimiento es el siguiente: la imagen se pasa a escala de grises, donde cada pixel tiene un valor de entre 0 y 255 (codificado en 8 bits), resultando $(b_7, b_6, b_5, b_4, b_3, b_2, b_1, b_0)$. Se realizan 8 planos con los valores 0 o 1 de cada pixel en cada plano, realizando posteriormente una conversión de cada uno a código de grises, que proporciona más información:

$$g_i = \begin{cases} b_i \oplus b_{i+1}, & 0 \leq i \leq m-2 \\ b_i, & i = m-1 \end{cases} \quad (42)$$

Donde \oplus es la operación XOR, b_i es el i-ésimo plano de bit e i es el i-ésimo plano expresado en escala de grises.

A continuación se hace uso de los cuatro planos superiores que contienen la mayor parte de la información estructural, y se obtiene con ellos un vector de entropías de cuatro componentes, una por cada plano.

La estructura de este descriptor es, por tanto, la de un vector con cuatro componentes, una por la entropía de cada plano de bits contemplados, de la forma $[E_1, E_2, E_3, E_4]$

5. *Luminancia: Luminancia media (LM)*

A partir del descriptor CLD de MPEG-7 se analiza el grado de continua de cada imagen, con el objetivo de identificar el nivel medio de la imagen. De esta forma se pueden diferenciar fácilmente aquellas imágenes luminosas de las que contienen tonos muy apagados u oscuros, con lo que se puede determinar la importancia de esta característica para cada usuario. Para ello, es necesario obtener una transformada DCT (Discrete Cosine Transform) sobre el espacio de color YCrCb de la imagen.

Este descriptor está compuesto por dos valores, incluyendo por una parte su valor medio, y por otra su varianza, que aporta información acerca de la dispersión de la primera, obteniéndose con ello un vector de dos dimensiones de la forma $[L_c, L_v]$

6. Crominancia: Saturación de HSV (SA)

En este caso se realiza un cambio de espacio de color para realizar el estudio del tono y la saturación de las imágenes elegidas, con el objetivo de determinar su impacto a la hora de su elección. Al igual que en el caso de la luminancia, este parámetro viene determinado por dos valores: por un lado el valor medio y por otro su varianza, de la forma $[S_m, S_v]$.

7. Crominancia: Variedad cromática (VC)

Descriptor generado a partir del DCD de MPEG7, se encarga de determinar la variación de color a lo largo de la imagen estudiada, obtenida a partir de los colores dominantes detectados. Viene dado a partir de tres valores: el número de colores dominantes, la varianza entre los diferentes tonos del mismo color, y la varianza entre los distintos colores (variedad intra e intercromática), de la forma $[N, VC_{intra}, VC_{inter}]$.

8. Crominancia: Coherencia espacial (CE)

Desarrollado a partir del DCD como en el caso anterior, su objetivo es determinar la continuidad o ausencia de la misma en cuanto al color de la imagen estudiada. Viene representado a partir de una única componente. Su extracción resulta análoga a la especificada para el descriptor DCD de MPEG-7 en [50].

9. Crominancia: Planaridad de color (PC)

Este descriptor representa la homogeneidad de cada color dominante en la imagen en términos de tono. Por ejemplo, una imagen en la que se repitan dentro de cada región de la imagen tonos muy similares tendrá una planaridad muy superior a la de una imagen abigarrada, en la que los colores se entremezclen de manera dispersa o caótica.

Se calcula de la siguiente manera: se divide la imagen en bloques cuadrados de píxeles, de tamaño proporcional al tamaño del lado menor de la imagen dividido por un coeficiente (que en nuestros tests ha tenido un valor de 8). Posteriormente se selecciona el color más frecuente de cada bloque y se calcula el porcentaje de píxeles que este contiene con un color similar (se ha estimado una diferencia de 20 dentro del sistema RGB). El valor final de la planaridad de color es el resultado del valor medio de esos porcentajes de píxeles para los distintos bloques.

Otros descriptores descartados

En el proceso de generación de descriptores relevantes, se obtuvieron otros descriptores de los que se esperaba obtener buenos resultados y que sirvieran para una clasificación más eficaz de las imágenes, y más acorde con la percepción sensorial del ojo humano. Sin embargo, las pruebas posteriores realizadas con usuario sirvieron para descartarlos, ya que no aportaban información relevante de las imágenes que resultara útil en la recomendación. De todos modos, en el futuro es posible seguir trabajando sobre ellos para poder llegar a incorporarlos con mejores resultados. Los descriptores descartados que se definieron con mayor claridad fueron dos:

1. Entropía general de la imagen

El estudio ha determinado que no es posible relacionar de forma coherente la entropía general de una imagen sin tratar con los intereses de cada usuario. Es decir, no puede inferirse los gustos de un usuario a partir de la entropía media de sus imágenes favoritas, por lo que su utilización se hace innecesaria en este caso.

2. Geometría de la imagen

El principal objetivo buscado con este análisis era el de determinar el eje de simetría (real o aproximado) de la imagen dada, para lo que se recurrió a diversos estudios previamente realizados. En los distintos artículos revisados, la geometría se analiza en imágenes muy concretas a partir de la segmentación de las mismas según regiones de interés, por lo que la problemática aumenta en el caso de imágenes naturales sin simetrías exactas.

Tal y como se explica en [56], existen dos métodos concretos para la detección de simetrías:

- Análisis del gradiente, óptimo para imágenes con un único objeto y simetrías bilaterales o rotacionales, por lo que su aplicación a las imágenes naturales no es recomendable.
- Análisis de la fase, aplicable en imágenes con más tipos de simetría. La principal característica de este método radica en que se encarga de detectar ejes de simetría unitarios y no genéricos, por lo que un estudio de la simetría genérica de la imagen sería imposible mediante su aplicación.

Los resultados obtenidos al aplicar el análisis de la fase (dado que es el método que puede proporcionar mejores resultados en imágenes cuyas simetrías no estén determinadas previamente) para la detección de simetrías de las imágenes tratadas no son relevantes, dado que se obtienen ejes de simetría unitarios, imposibilitando con ello su utilización para la detección de simetrías genéricas, tal y como se perseguía.

5.4 Modelo de extracción de características y anotación de vídeos para la recomendación

5.4.1 Caracterización de atributos fílmicos

Como se afirma en [57], donde se realiza un análisis audiovisual de dos textos fílmicos específicos, el uso de determinadas herramientas de la sintaxis audiovisual caracteriza estilos formales que suelen ser agrupados y etiquetados a más alto nivel, de tal forma que algunos patrones estilísticos dan lugar a determinadas corrientes o movimientos cinematográficos que ocasionan una especial atracción o repulsión por parte de diferentes tipologías de espectadores. Resulta evidente, por lo tanto, que el estudio de características estructurales y morfológicas de la imagen y de la sintaxis audiovisual permite crear *clusters* estilísticos, agrupando aquellas obras que pertenecen a corrientes determinadas, con lo cual es posible clasificar automáticamente las obras audiovisuales

en categorías semánticas utilizadas en otro tipo de recomendadores, como el país de procedencia, la época o su clasificación dentro de la historia del cine. Un análisis de este tipo permite incluso llegar a niveles de clasificación más finos, haciéndose posible discriminar obras de determinados directores, de un estilo muy homogéneo, o encontrar concomitancias o diferencias muy significativas entre obras diversas o relacionadas.

Como afirma Domenec Font [46], la creciente heterogeneización y transnacionalización del audiovisual ha ocasionado, en los últimos tiempos, que determinadas corrientes estilísticas, tradicionalmente asociadas a una determinada cinematografía, invadan las formas de creación audiovisual de otros lugares, con lo cual las etiquetas tradicionales pierden su validez, al hibridarse influencias procedentes de épocas y países muy distintos. Existen, por esta razón, determinadas obras ancladas en la tradición fílmica de un país y una época que, sin embargo, se asemejan formalmente en mayor medida a otras corrientes estilísticas, lo que provoca que ciertos espectadores se vean desplazados respecto a sus previsiones iniciales. Un ejemplo evidente puede ser el de ciertas películas de Gus van Sant, cineasta estadounidense contemporáneo habitualmente clasificado con la etiqueta *indie*. Si bien algunas películas de este cineasta sí se adaptan a los patrones de este tipo de cine, algunas, como *Gerry* (2002), *Elephant* (2003) o *Last days* (2005), están mucho más próximas al estilo de cineastas europeos de la modernidad como Bela Tarr y Chantal Akerman, como han señalado importantes analistas como Jonathan Rosenbaum [58]. Un análisis de características formales, como la duración media de los planos, la homogeneidad de la textura de las imágenes, la intensidad y complejidad de los movimientos de cámara, o la abundancia de silencios, agruparía estas obras junto con las de los cineastas europeos citados por Rosenbaum, logrando, de cara a la recomendación, una eficiencia superior que la de los recomendadores semánticos, como se demuestra en el capítulo 6 de la tesis. Otro ejemplo intuitivo sería el de análisis de la obra de determinados cineastas, como el danés Carl Theodor Dreyer, que los analistas suelen asociar con la tasa de planos empleados [48], ya que se aprecia a lo largo de su obra, desde *La pasión de Juana de Arco* (1929) hasta *Gertrud* (1964), una reducción lineal del número de planos empleados (debido al aumento en la duración de estos), con lo que un análisis formal como el propuesto sería capaz de extraer conclusiones de esta naturaleza, de gran utilidad para la recomendación. Por lo tanto, una recomendación formal flexible y dinámica como la que se propone, en determinados ámbitos puede llegar a ser más precisa para la descripción de contenidos audiovisuales que la habitual categorización de conceptos.

5.4.2 Modelo de descriptores fílmicos para la recomendación

Como afirman Aumont *et al.* en su *Estética del cine* [41], el discurso fílmico se articula como un esquema de representación tanto visual como sonora, por lo que los descriptores obtenidos deberían cubrir ambas ramas. Dado que el trabajo de esta tesis se centra en la parte visual, se centra aquí el esfuerzo en una definición más exhaustiva de descriptores; sin embargo, debido a su especial relevancia y su utilidad en la descripción de cierto tipo de obras audiovisuales, también se consideran algunos descriptores de audio de alto nivel.

En el plano visual, Aumont basa su análisis del lenguaje cinematográfico (que a su vez extrapola a todo el lenguaje audiovisual) en tres pilares: la imagen, el movimiento y el

montaje. En esta definición se basa la clasificación de descriptores fílmicos visuales, que junto a los descriptores de audio conforman los cuatro tipos de descriptores analizados en la propuesta de esta tesis:

- a) **Imagen.** Continúa Aumont diciendo que la concepción de la imagen a partir de la definición de “cuadro” es la que une el cine con otras artes como la pintura o la fotografía, mientras que el movimiento y el montaje aporta su especificidad. La imagen, por lo tanto, se puede describir formalmente en los términos empleados en el apartado anterior, con algunas particularidades técnicas ocasionadas por la dificultad del tratamiento de vídeo respecto al tratamiento de imágenes.

En primer lugar, es importante tener en cuenta la presencia de la luz y del color, para lo que se introducen descriptores de **luminosidad** de la obra, a través de sus valores de media y varianza, y de color, a través de los valores de **saturación** y de **variedad cromática**. Además, presenta un especial interés para la recomendación el hecho de que una obra esté hecha en blanco y negro o en color, ya que en muchas ocasiones esto va más allá de la percepción implícita de un usuario, ya que este expresa un rechazo explícito a ciertas obras por estos motivos, basados en prejuicios. Por esta razón se incluye un descriptor extra, **tasa de blanco y negro**, que sería redundante en términos de percepción, pero que es necesario para caracterizar las preferencias de aquellos que utilizan esta característica como restricción para su consumo de obras audiovisuales.

Por último, existen otras peculiaridades que cobran especial relevancia en la obra cinematográfica, como son el ángulo de la cámara y la escala de plano. En cuanto al ángulo de la cámara, todavía no se encuentran en el estado del arte aproximaciones suficientemente finas como para caracterizar este atributo y, en cualquier caso, la pérdida de eficiencia tendría otras contraindicaciones, por lo que se ha descartado representar esta cualidad. En cuanto a la escala de planos, esto resulta mucho más representativo de diferentes estilos fílmicos, dado que, por ejemplo, numerosos cineastas reniegan del primer plano o minimizan su uso, priorizando el plano general, o en determinadas gramáticas cinematográficas se utilizan escalas muy determinadas (por ejemplo, el plano americano es muy habitual en el cine clásico de Hollywood, mientras que el primer plano es muy utilizado en películas recientes de acción, o en melodramas efectistas de tradición hollywoodiense). El reconocimiento de “cromos” y de diferentes escalas de planos es un tema actualmente tratado en el estado del arte, pero su eficiencia todavía no garantiza un nivel de acierto que permita caracterizar adecuadamente esta cualidad del audiovisual. Sin embargo, se ha encontrado una manera de aproximar la medida de presencia de planos más cortos, como el “plano detalle” o el “primer plano”, o planos más largos, como el “plano general” o el “gran plano general”, que es la medida de la cantidad de información presente en la imagen. Los planos lejanos, habitualmente, aportan una mayor cantidad de información a la imagen (razón por la cual su duración suele ser mayor), mientras que los planos más cercanos suelen contener menos información (que también hace que los cineastas les ajusten una duración menor, ya que el ojo humano necesita menos tiempo para captar menos información). La medida clásica de la cantidad de información en el tratamiento de la señal es la **entropía**, por lo que se incluye como un descriptor importante de tipo visual.

b) **Movimiento.** En cuanto al movimiento, Aumont diferencia entre el movimiento interno al cuadro, que puede venir provocado por el movimiento de los personajes o la traslación de objetos, y el movimiento de cámara, respecto al cual establece, a su vez, tres tipos de movimientos, panorámica, travelling y zoom. Por lo tanto, resulta fundamental diferenciar entre el movimiento de cuadro o de cámara, ya que modifican radicalmente la percepción estética de la obra. Sin embargo, la tipología de movimientos de cámara es algo más transparente a un usuario común. Este usuario, aunque no rechaza o se sienta atraído necesariamente por una determinada tipología de movimientos de cámara, sí resulta más sensible a la intensidad y a la complejidad de estos movimientos. Por ejemplo, una película rodada cámara en mano (lo cual provoca el rechazo de algunos espectadores) presentaría altos valores tanto de intensidad como de complejidad de los movimientos de cámara, mientras que una película actual realizada en un gran estudio llevaría asociados grandes movimientos de cámara (en grúas o *steadycam*), que por lo tanto también darían un alto valor de intensidad, pero que serían mucho más estables y regulares, es decir, con valores más bajos de complejidad. Por su parte, los valores de intensidad permiten diferenciar estilos muy diferentes, como aquellos más austeros de los más efectistas, o los más clásicos de los más actuales. Estas características de intensidad y complejidad, además, están recogidas en la definición de descriptores de movimiento del estándar MPEG-7 [50]. Por lo tanto, el movimiento de los vídeos se caracteriza en el modelo propuesto en esta tesis a través de la medida de la **intensidad media del movimiento de cámara**, la **varianza de la intensidad del movimiento de cámara**, la **media de la complejidad del movimiento de cámara**, la **varianza de la complejidad del movimiento de cámara**, la **intensidad media del movimiento de cuadro**, la **varianza de la intensidad del movimiento de cuadro**, la **media de la complejidad del movimiento de cuadro**, y la **varianza de la complejidad del movimiento de cuadro**.

c) **Montaje.** Refiriéndose al montaje, Aumont habla de tres funciones principales: función sintáctica, función semántica y función rítmica.

La función sintáctica es la que resulta más fácilmente cuantificable, ya que simplemente es aquella mediante la cual el montaje, empleando los elementos cinematográficos disponibles, une los distintos fragmentos de materia filmada. Estos elementos son tradicionalmente tres, el corte, el fundido y el encadenado, que equivalen a los signos de puntuación del lenguaje cinematográfico. Todos los planos de que consta un material audiovisual se unen mediante algunos de estos elementos, por lo que, finalmente, para la caracterización del vídeo se procede a la **detección de planos**, **detección de fundidos** y **detección de encadenados**. El número de cortes sería redundante con los anteriores, por lo que solo se cuantifica como paso intermedio para la medida de los planos.

Del mismo modo, también es necesario cuantificar la **duración de los planos** para cubrir la función rítmica del montaje. Esta función rítmica, además del descriptor anterior, queda cuantificada representando la manera en que se distribuyen los elementos sintácticos a lo largo del contenido audiovisual, es decir, estableciendo medidas de dispersión como la **varianza asociada a cada uno de los elementos sintácticos**.

La cuantificación de la función semántica, por su parte, debido a las características y restricciones impuestas en los objetivos de la tesis, queda en su mayor parte fuera de las posibilidades de este estudio. Sin embargo, se ha podido realizar una aproximación a través que la narratividad que suele ir asociada a las diferentes escenas y secuencias que estructuran la mayor parte de los contenidos audiovisuales¹². La segmentación de secuencias queda totalmente fuera del alcance de la tesis, ya que requiere el uso de elementos semánticos, pero sí es posible realizar una aproximación de la **segmentación de escenas**, ya que estas se corresponden con la representación de imágenes en un mismo espacio escénico, que puede ser deducido a través de características de bajo nivel de las imágenes, como la luz, el color o las texturas, teniendo en cuenta algunos conceptos de construcción sintáctica audiovisual, tal y como se explica en el apartado 5.4.6 (Capa visual. Descriptores de alto nivel), en la definición formal del atributo correspondiente a la segmentación de escenas. Además, el modelo propuesto permite categorizar las obras fílmicas a partir de los parámetros de las dos teorías de montaje mayoritarias, las correspondientes a Eisenstein [59] y a André Bazin [60], ya que la primero se focaliza (además del montaje ideológico, inabordable según los parámetros tratados) en la función rítmica, y la segunda en la función sintáctica.

- d) **Audio.** Finalmente, y a pesar de que el foco de la tesis está fijado en la parte visual, se han considerado algunos descriptores de alto nivel relacionados con el audio, debido a su especial interés de cara a la recomendación. A modo de experimento, se consideró establecer los parámetros más representativos de los diversos tipos de narrativa audiovisual empleada en la actualidad y a lo largo del último siglo. Por estas razones, se segmenta el tiempo total de audio entre **tiempo de habla**, que corresponde a aquel en el que existe habla (ya sea por parte de los personajes o de una *voz over*), aquel en el que existe **silencio**, o aquel en el que se produce algún **otro efecto sonoro**, ya se trate de banda sonora o sonidos extradiegéticos, o de efectos sonoros correspondientes a la acción.

Hubiera sido interesante haber podido distinguir entre estos dos últimos tipos, sonido y banda sonora mayoritariamente, pero después de realizar pruebas con la últimas herramientas de tratamiento de audio empleadas mayoritariamente en el estado del arte, se determinó que el funcionamiento no era suficientemente preciso.

Por último, es necesario señalar que la caracterización de ciertos descriptores, especialmente entre los de montaje y movimiento, a partir de los valores de media y varianza fue decidido después de comprobar, en vídeos de una duración mínima (todos los vídeos considerados en el corpus tienen una duración superior a la hora), que los valores de esos descriptores seguían distribuciones *gaussianas* o que se podían

¹² La mayor parte de la producción audiovisual actual y del pasado, como afirma Aumont, tiene carácter narrativo, aunque se trate de una narratividad suave. La gran mayoría de la producción audiovisual tiene carácter figurativo, lo cual siempre lleva implícito un cierto concepto de narratividad, por lo que el cine no narrativo solo pertenecen algunas obras muy específicas de cineastas underground o de vanguardia como Stan Brackage, Michael Snow o Werner Nekes, quienes tienen como propósito, precisamente, articular reflexiones críticas acerca de conceptos clave del lenguaje audiovisual, como la ficción, el dispositivo, etc.

aproximar por *gaussianas*. Esto se determinó aplicando pruebas de contraste de hipótesis de normalidad, siguiendo el algoritmo de Kolmogorov-Smirnov, que se describe en el apartado de pruebas y validación de este mismo capítulo. Por lo tanto, una caracterización completa es posible mediante el uso de los parámetros media y varianza.

5.4.3 Modelo de extracción automática

La extracción automática de los metadatos definidos se articula a través de varias capas de procesado, una primera de descriptores visuales de bajo nivel, una segunda de descriptores visuales de alto nivel, y una última que utiliza los resultados de los descriptores previos, alto y bajo nivel, para generar, a través de las herramientas adecuadas, los valores finales de los descriptores definidos.

El modelo se desarrolla a partir del esquema de la Fig. 31

En primer lugar, se obtienen dos características de bajo nivel sobre todos los *frames* que componen el vídeo: gradiente y vectores de movimiento. Además, sobre algunos *frames* candidatos se aplican otras características de bajo nivel: histogramas de color y medidas de luminancia. A partir de estas medidas, se realizan modelados más complejos y se toman decisiones a través de algoritmos que también tienen en cuenta peculiaridades de la narrativa audiovisual. Estas medidas permiten detectar los movimientos de cuadro y de cámara, y segmentar el vídeo en planos a través de la detección de cortes, fundidos y encadenados. De cada uno de los planos se extrae uno de sus *frames* más representativos, o *key-frame*, sobre el cual se vuelven a realizar operaciones de bajo nivel para obtener los descriptores de imagen. Además, los *key-frames* también se emplean para realizar una aproximación a la detección de escenas. Por su parte, también se modelan las medidas definitivas de los descriptores de montaje y de movimiento a partir de los descriptores visuales de alto nivel previamente obtenidos.

De esta manera, se caracteriza un vídeo según el modelado estético diseñado, obteniendo los descriptores necesarios para la etapa final de recomendación.

Los siguientes apartados formalizan los valores de los metadatos finales y explican las estrategias empleadas para generar los descriptores visuales de bajo y alto nivel.

5.4.4 Capa de metadatos

La Tabla 18 agrupa los 26 descriptores finales definidos, clasificándolos en los tipos ya definidos en esta tesis, y asociándolos con el acrónimo que se les ha asignado para facilitar su identificación y con el lugar que les correspondería en la jerarquía de metadatos visuales establecida en el estándar MPEG-7.

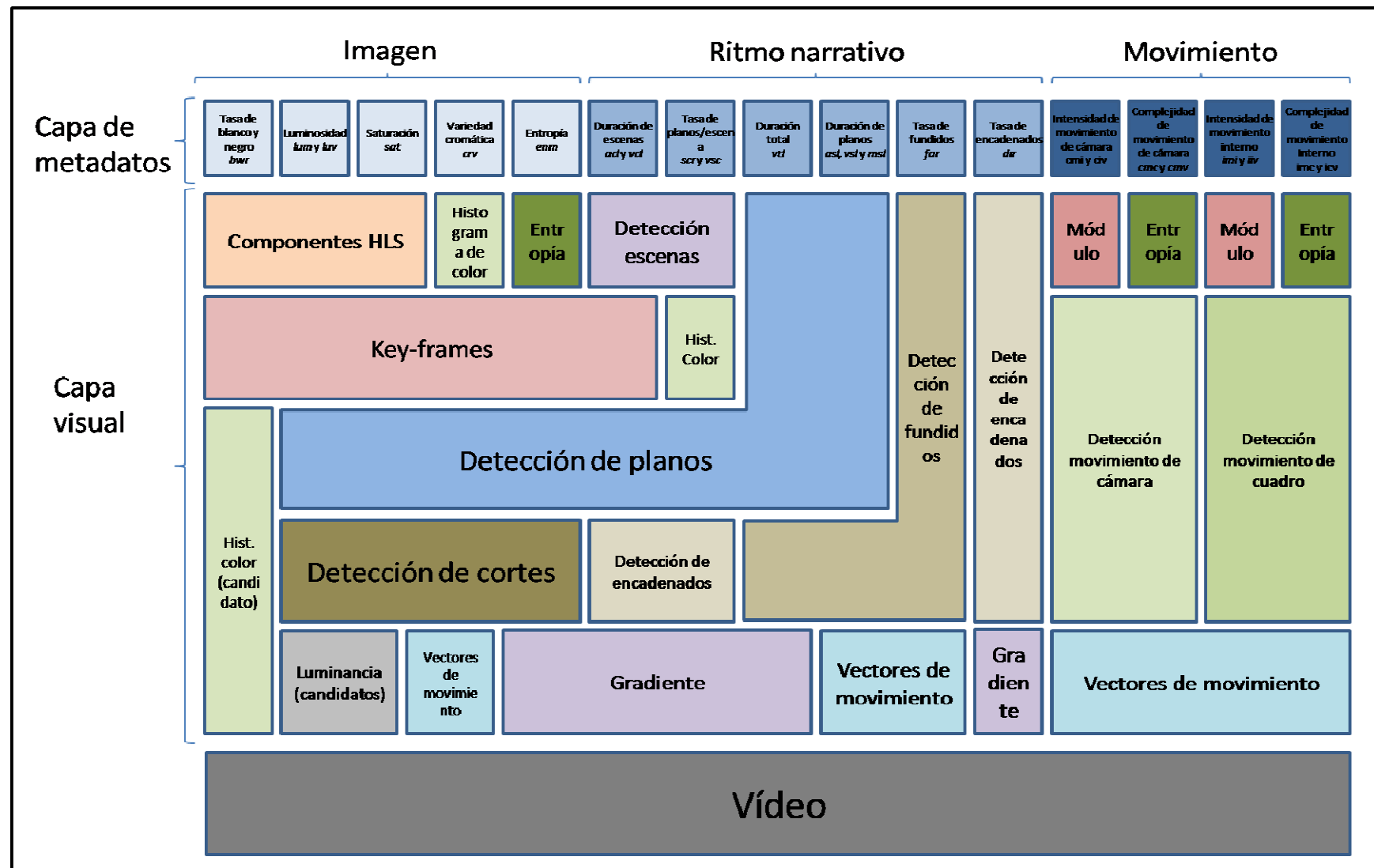


Fig. 31. Modelo de descripción formal y estética de vídeos

Nº	Tipo	Descriptor	Acrónimo	Jerarquía MPEG-7
1	Imagen	Tasa de blanco y negro	BWR	Structural → Visual → Color
2	Imagen	Luminosidad media	LUM	Structural → Visual → Color
3	Imagen	Varianza de la luminosidad	LUV	Structural → Visual → Color
4	Imagen	Saturación	SAT	Structural → Visual → Color
5	Imagen	Variedad cromática	CRV	Structural → Visual → Color
6	Imagen	Entropía media	ENM	Structural → Visual → Texture
7	Montaje	Duración media de planos	ASL	Structural → Visual → Narrative
8	Montaje	Heterogeneidad de duración de planos	VSL	Structural → Visual → Narrative
9	Montaje	Mediana de duración de planos	MSL	Structural → Visual → Narrative
10	Montaje	Duración total	VTL	Structural → Visual → Narrative
11	Montaje	Duración media de escenas	ACL	Structural → Visual → Narrative
12	Montaje	Varianza de la duración de escenas	VCL	Structural → Visual → Narrative
13	Montaje	Tasa de planos por escena	SCR	Structural → Visual → Narrative
14	Montaje	Varianza de la tasa de planos por escena	VSC	Structural → Visual → Narrative
15	Montaje	Tasa de fundidos	FAR	Structural → Visual → Narrative
16	Montaje	Tasa de encadenados	DIR	Structural → Visual → Narrative
17	Movim	Intensidad media del movimiento de cámara	CMI	Structural → Visual → Motion
18	Movim	Complejidad media del movim de cámara	CMC	Structural → Visual → Motion
19	Movim	Intensidad media del movimiento interno	IMI	Structural → Visual → Motion
20	Movim	Complejidad media del movimiento interno	IMC	Structural → Visual → Motion
21	Movim	Varianza de la intensidad mov cámara	CIV	Structural → Visual → Motion
22	Movim	Varianza de la complejidad mov cámara	CMV	Structural → Visual → Motion
23	Movim	Varianza de la intensidad del mov interno	IIV	Structural → Visual → Motion
24	Movim	Varianza de la complejidad del mov interno	ICV	Structural → Visual → Motion
25	Audio	Tasa de habla	SPR	Structural → Audio
26	Audio	Tasa de silencio	SIR	Structural → Audio

Tabla 18. Descriptores incluidos en el modelo

A continuación, se formaliza el valor de cada uno de estos descriptores. Es importante reseñar que se ha primado la eficiencia computacional dentro de las necesidades del sistema de extracción automática.

IMAGEN: Tasa de blanco y negro (BWR)

Porcentaje del tiempo total en que la imagen es en blanco y negro, *Black and White Rate* (BWR):

$$BWR = \frac{T_{BN}}{T} \quad (43)$$

Donde T_{BN} es el tiempo de duración de la parte en blanco y negro y T la duración total del vídeo.

Para su extracción no es necesario el procesado de todos los *frames* que componen el vídeo, ya que su objetivo es dar una idea cualitativa del contenido percibido por el usuario. Es suficiente con comprobar la presencia de color en los *key-frames* extraídos. Esta idea se extrapola a todos los descriptores del tipo imagen.

IMAGEN: Luminosidad media (LUM)

Analizando cada imagen según sus coordenadas HSL (Hue-Saturation-Lightness), se calcula la luminancia media (L_m) de los K key-frames que componen el vídeo. El valor final será el valor medio de los valores de luminancia media de cada key-frame, siendo esta, a su vez, definida como el valor medio de los valores de intensidad de cada uno de los píxeles que forman la imagen.

$$LUM = L_{AVE} = \frac{1}{K} \sum_{k=1}^K L_m \quad (44)$$

IMAGEN: Varianza de la luminosidad (LUV)

Se obtiene calculando la varianza de los valores de luminancia media (L_m) de los K *key-frames* escogidos, empleando los valores medios (L_{AVE}) obtenidos con el descriptor LUM.

$$LUV = \sigma_S^2 = \frac{1}{K} \sum_{k=1}^K (L_m - L_{AVE})^2 \quad (45)$$

IMAGEN: Saturación (SAT)

Analizando cada imagen según sus coordenadas HSL, se calcula la media de la saturación media (S_m) de los K *key-frames* escogidos, donde la saturación media de un *frame* es el valor medio de los valores de saturación de cada uno de los píxeles que componen la imagen.

$$SAT = S_{AVE} = \frac{1}{K} \sum_{k=1}^K S_m \quad (46)$$

IMAGEN: Variedad cromática (CRV)

Se define la variedad cromática de un *frame* i (VC_i) como la varianza del histograma de *bins* de colores obtenido de cada imagen. N es el número de *bins*. La medida final será la media de los valores de variedad cromática de todos los *key-frames* presentes en el vídeo.

$$VC_i = \frac{1}{K} \sum_{k=1}^K N \cdot (B_n - B_{AVE})^2 \quad (47)$$

$$VC_{AVE} = \frac{1}{K} \sum_{k=1}^K VC_i \quad (48)$$

IMAGEN: Entropía (ENM)

Se calcula el valor de la entropía de cada *key-frame*, empleando para ello los valores de intensidad de la luminancia, sobre sus N posibles valores cuantificados. Finalmente, se obtiene la media de esos resultados para los K *key-frames* que componen el vídeo.

$$ENM = \frac{1}{K} \sum_{k=1}^K \sum_i^N -P(L_{ki}) \cdot \log_2 L_{ki} \quad (49)$$

MONTAJE: Duración media de planos (ASL)

Se calcula obteniendo el valor medio de la duración de los S_N planos de que componen el vídeo. Estos son previamente obtenidos mediante los algoritmos de detección de cortes, fundidos y encadenados:

$$ASL = S_{AVE} = \frac{1}{N} \sum_{n=1}^N S_n \quad (50)$$

Donde N es el número total de planos del vídeo.

MONTAJE: Heterogeneidad de la duración de planos (VSL)

Se calcula obteniendo la desviación típica de la duración de todos los planos del vídeo:

$$VSL = \sigma_s = \sqrt{\frac{1}{N} \sum_{n=1}^N (S_n - S_{AVE})^2} \quad (51)$$

Donde N es el número total de planos del vídeo.

MONTAJE: Duración mediana de planos (MSL)

Se obtiene seleccionando la mediana del valor de duración de los planos del vídeo, es decir, aquel situado en el punto medio del rango de duraciones de planos.

$$MSL = Mediana = x_{i1} + \left(\frac{\left(\frac{N_M}{2} \right) - N_{i-1}}{f_i} \right) \cdot (x_{i2} - x_{i1}) \quad (52)$$

MONTAJE: Duración total (VTL)

La duración total del vídeo se obtiene de la suma de las duraciones de los N planos que lo componen.

$$VTL = \sum_{n=1}^N S_n \quad (53)$$

MONTAJE: Duración media de escenas (ACL)

Se obtiene a partir de la detección de escenas, calculando la media de la duración de todas las escenas del vídeo:

$$ACL = C_{AVE} = \frac{1}{M} \sum_{m=1}^M C_m \quad (54)$$

Donde M es el número total de escenas del vídeo.

MONTAJE: Varianza de la duración de escenas (VCL)

Se obtiene del cálculo de la varianza de la duración de las escenas detectadas en el vídeo:

$$VCL = \sigma_S^2 = \frac{1}{M} \sum_{m=1}^M (C_m - C_{AVE})^2 \quad (55)$$

Donde M es el número total de escenas del vídeo.

MONTAJE: Tasa de planos por escena (SCR)

Se obtiene calculando el valor medio del número de planos N_m de cada escena m , a partir de la extracción automática de planos y escenas.

$$SR_{AVE} = \frac{1}{M} \sum_{m=1}^M N_m \quad (56)$$

MONTAJE: Varianza de la tasa de planos por escena (VSC)

Se obtiene calculando la varianza del número de planos por escena:

$$\sigma_{SR}^2 = \frac{1}{M} \sum_{m=1}^M (N_m - SR_{AVE})^2 \quad (57)$$

MONTAJE: Tasa de fundidos (FAR)

La Tasa de fundidos, o Fade Rate, se obtiene dividiendo el número de fundidos detectados por el número total de transiciones del vídeo:

$$FAR = \frac{F}{N-1} \quad (58)$$

Donde F es el número de fundidos del vídeo y N es el número total de planos del vídeo.

MONTAJE: Tasa de encadenados (DIR)

La Tasa de encadenador, o Dissolve Rate, se obtiene dividiendo el número de encadenados detectados por el número total de transiciones del vídeo:

$$DR = \frac{D}{N-1} \quad (59)$$

Donde D es el número de encadenados del vídeo y N es el número total de planos del vídeo.

MOVIMIENTO: Intensidad media del movimiento de cámara (CMI)

En primer lugar es necesario detectar si existe movimiento de cámara. En ese caso, se aplica la fórmula de intensidad para aquellos vectores de movimiento asociados al movimiento de la cámara, es decir, aquellos que se sitúan en el borde del cuadro.

Como cada macrobloque (i,j) tiene un vector de movimiento $(x_{i,j}, y_{i,j})$

$$MI_n = \frac{1}{Q} \sum_{w=1}^Q \sum_{i=1}^N \sqrt{x_{i,j}^2 + y_{i,j}^2} \quad (60)$$

$$CMI = \frac{1}{N} \sum_{n=1}^N MI_n \quad (61)$$

Q es el número de frames que se utilizan de cada plano y N es el número de vectores de movimiento empleados.

MOVIMIENTO: Complejidad media del movimiento de cámara (CMC)

En primer lugar es necesario detectar si existe movimiento de cámara. En ese caso, se calcula un histograma de orientaciones de los vectores de movimiento asociados al movimiento de la cámara (situados en el borde del cuadro), y luego se calcula su entropía:

$$MC_n = \frac{-1}{Q} \sum_{w=1}^Q \sum_{n=1}^N h(n) \log(h(n)) \quad (62)$$

$$CMC = \frac{1}{N} \sum_{n=1}^N MC_n \quad (63)$$

MOVIMIENTO: Intensidad media del movimiento de cuadro (IMI)

Después de haber obtenido la intensidad del movimiento debido únicamente al movimiento de la cámara, se halla el movimiento general de la imagen en el *frame* n (MV_n), debido tanto a la cámara como a los objetos que se mueven en el cuadro. Esto se realiza de forma análoga a la de las expresiones (60) y (61), pero considerando todos los vectores de movimiento de referencia dentro del cuadro, y no solamente los del borde. Haciendo la sustracción, se obtiene el valor de la intensidad correspondiente al movimiento de cuadro.

$$IMI = \frac{1}{N} \sum_{n=1}^N MV_n - MI_n \quad (64)$$

MOVIMIENTO: Complejidad media del movimiento de cuadro (IMC)

La definición es análoga a la del descriptor *CMC*. La diferencia está en la construcción del histograma. En lugar de considerar los vectores de referencia del borde del cuadro, se consideran como referencia los vectores internos. Una vez obtenido su histograma de orientaciones, se restan los valores del histograma del movimiento de cámara, quedando reflejados, por lo tanto, los movimientos que corresponden a los personajes y objetos del cuadro. Sobre los valores obtenidos (MS_n), se calcula la media aritmética, cuyo valor corresponderá con el del descriptor *IMC*.

MOVIMIENTO: Varianza de la intensidad del movimiento de cámara (CIV)

Se calcula la varianza de los valores obtenidos de MI_n .

$$CIV = \frac{1}{N} \sum_{n=1}^N (MI_n - CMI)^2 \quad (65)$$

MOVIMIENTO: Varianza de la complejidad del movimiento de cámara (CMV)

Se calcula la varianza de los valores obtenidos de MC_n .

$$CMV = \frac{1}{N} \sum_{n=1}^N (MC_n - CMC)^2 \quad (66)$$

MOVIMIENTO: Varianza de la intensidad del movimiento de cuadro (IIV)

Se calcula la varianza de los valores obtenidos de $MV_n - MC_n$.

$$IIV = \frac{1}{N} \sum_{n=1}^N ((MV_n - MC_n) - IMC)^2 \quad (67)$$

MOVIMIENTO: Varianza de la complejidad del movimiento de cuadro (ICV)

Se calcula la varianza de los valores obtenidos de $MS_n - MI_n$.

$$ICV = \frac{1}{N} \sum_{n=1}^N ((MS_n - MI_n) - IMI)^2 \quad (68)$$

AUDIO: Tasa de habla (SPR)

Se calcula la tasa de habla dividiendo el tiempo total en el que se produce habla (de algún personaje o de una *voz over*) por el tiempo total de duración del vídeo.

$$SPR = \frac{T_{SP}}{T} \quad (69)$$

AUDIO: Tasa de silencio (SIR)

Se calcula la tasa de silencio dividiendo el tiempo de silencio que se detecta por el tiempo total de duración del vídeo.

$$SIR = \frac{T_{SI}}{T} \quad (70)$$

5.4.5 Capa visual. Descriptores de bajo nivel

Gradiente

El gradiente es uno de los descriptores que, como ya se ha reseñado, se obtiene sobre cada uno de los *frames* que componen el vídeo, ya que su valor es una de las bases para la obtención de descriptores de más alto nivel.

El gradiente da una idea cualitativa (ya que depende de los valores de umbralización) del número de bordes de una imagen, lo cual es útil para deducir distintas características del vídeo. Esta presencia de bordes detectados no responde sólo a la cantidad de objetos que aparecen en una imagen (por ejemplo, normalmente en un plano general habrá más bordes que en un primer plano), sino también a la nitidez de la imagen, ya que esto condiciona la detección.

Para la obtención del gradiente se realiza una doble operación. En primer lugar, se emplea un filtro de Canny [61] para la binarización de la imagen. La decisión de utilizar un operador de segundo orden se debe a que la eficiencia de los operadores de primer orden no es suficiente para los resultados que necesitan las capas superiores del modelo de descriptores. A continuación, se utiliza el algoritmo de detección de bordes de Suzuki et al. [62] para obtener la cuantificación numérica definitiva del número de bordes, que permite utilizar directamente el resultado para realizar operaciones de más alto nivel.

Vectores de movimiento

La realización de las medidas de intensidad y orientación de los vectores de movimiento responde a la necesidad de caracterizar la cantidad y el tipo de movimiento de un vídeo en función de la percepción del usuario, para lo que se diferencia entre movimiento de cámara y movimiento de cuadro, como se explica en la sección de descriptores visuales de alto nivel.

Para realizar la medida se utiliza un método que describe las coordenadas del movimiento de una serie de puntos previamente fijados. Esto genera un módulo y un ángulo del movimiento con respecto a su anterior posición, lo cual define la intensidad y la complejidad del movimiento de cada uno de los vectores.

Este algoritmo está basado en el método de Lucas-Kanade [63] para la detección del movimiento mediante pirámides, que aun siendo una aproximación de primer orden para la obtención del flujo óptico, es suficiente para una aplicación en la simplemente es necesario cuantificar en términos relativos el movimiento de ciertos puntos seleccionados de la imagen.

Componentes HLS e histograma de color

Entre las medidas de luminancia y color empleadas para caracterizar las imágenes están la extracción de las componentes HLS de los píxeles que componen la imagen, y la construcción de histogramas de color que se emplean para comparar dos *frames* del vídeo.

A continuación se describe la construcción del histograma de color, empleando las cuantificaciones de *bins* definidas por [64].

El espacio de color escogido para generar los histogramas de color, del mismo modo que para el resto de medidas, ha sido el HSL (Hue-Saturation-Lightness) porque la sensibilidad del ojo humano es mucho más cercana a éste que al RGB (Red-Green-Blue) tradicional [50].

El histograma consta de 12 *bins* para el tono (Hue), 4 *bins* para la saturación (Saturation) y 5 *bins* para la luminosidad (Lightness). Es decir, existe un valor del histograma color para los píxeles con un valor de *bin* de tono 1, saturación 1 y luminosidad 1, otro para los píxeles con un valor de *bin* de tono 1, saturación 1 y luminosidad 2, y así sucesivamente hasta completar todas las combinaciones.

De esta manera, el tamaño total del histograma de color es de 240 *bins* (12*5*4). La ecuación que relaciona el bin C en el que está un pixel, con sus valores cuantificados H, S y L de tono, saturación y luminosidad, viene dado por (70).

$$C = 20 \cdot (H - 1) + 4 \cdot (S - 1) + L \quad (71)$$

A su vez, para obtener los valores de tono, saturación y luminosidad con respecto a los RGB se utilizan las siguientes relaciones.

Siendo $M = \text{MAX}(R, G, B)$ y $m = \text{MIN}(R, G, B)$

$$H = \begin{cases} 0, & \text{si } M = m \\ (60 \times \frac{G-B}{M-m} + 360) \bmod 360, & \text{si } M=R \\ 60 \times \frac{B-R}{M-m} + 120, & \text{si } M=G \\ 60 \times \frac{R-G}{M-m} + 240, & \text{si } M=B \end{cases} \quad (72)$$

$$L = \frac{1}{2}(M + m) \quad (73)$$

$$S = \begin{cases} 0, & \text{si } M = m \\ \frac{M-m}{2L}, & \text{si } L \leq \frac{1}{2} \\ \frac{M-m}{2(1-L)}, & \text{si } L > \frac{1}{2} \end{cases} \quad (74)$$

Una vez obtenidos los histogramas, el siguiente paso para realizar la comparación es obtener la media, la desviación típica y el tercer momento del histograma de color. Se define de la siguiente manera:

Media del histograma i

$$\mu_i = \frac{1}{N} \sum_{j=1}^M P_{ij} \quad (75)$$

Desviación típica del histograma i

$$\sigma_i = \sqrt{\frac{1}{N} \sum_{j=1}^M (P_{ij} - \mu_i)^2} \quad (76)$$

Tercer momento (o asimetría) del histograma i

$$\gamma_i = \sqrt[3]{\frac{1}{N} \sum_{j=1}^M (P_{ij} - \mu_i)^3} \quad (77)$$

Donde N es el número de elementos de la imagen, M es el número de *bins* del histograma de color y P_{ij} es el valor del bin j del histograma i .

Finalmente, la relación empleada para determinar la distancia entre histogramas a partir de estas métricas viene dada por (78).

$$d(H, I) = \sum_{i=1}^n (W_{i1} |\mu_{iH} - \mu_{iI}| + W_{i2} |\sigma_{iH} - \sigma_{iI}| + W_{i3} |\gamma_{iH} - \gamma_{iI}|) \quad (78)$$

5.4.6 Capa visual. Descriptores de alto nivel

Detección de movimiento de cámara y de cuadro

Para la obtención de los metadatos definitivos relativos al movimiento tanto de cámara como de cuadro se emplean las medidas de bajo nivel de los vectores de movimiento, como ya ha sido explicado. Sin embargo, es necesario distinguir entre el movimiento que responde al movimiento de la cámara y el que corresponde con el cambio de posición de objetos y posiciones dentro de un cuadro considerado fijo.

Con esta finalidad, se obtienen las medidas de dos grupos de puntos fijados previamente.

El primer grupo está formado por 8 puntos exteriores (en verde en la imagen), de tal manera que se pueda inferir el movimiento de la cámara. Un movimiento de cámara implica el movimiento de todos los puntos del cuadro, por lo que es necesario situar los puntos lo más alejados posible y lo más lejos posible de los focos habituales de acción. Al situar los puntos en el borde del cuadro se reduce drásticamente la posibilidad de que el movimiento de todos ellos corresponda con el movimiento de uno o varios personajes u objetos simultáneamente, incluso en los casos de primeros planos, que serían conflictivos en el caso de que los puntos se situaran en el interior de la imagen.

El segundo grupo, por su parte, está formado por 9 puntos interiores (en morado en la imagen), dispuestos de forma equidistante para cubrir el conjunto de la imagen de la forma más representativa posible. Estos puntos se utilizan para medir el movimiento de

cuadro, es decir, el de los objetos y personajes que aparecen en la imagen, sin considerar el movimiento de cámara, para lo que es necesario restar al movimiento general aquel que corresponde únicamente con el movimiento de cámara, como se define en el apartado “Capa de metadatos” (5.4.4) para los descriptores correspondientes.

Por lo tanto, el mayor problema es decidir cuándo se produce movimiento de cámara. Para ello, se tiene en cuenta una serie de consideraciones:

- La disparidad de los ángulos de los puntos tomados como referencia debe ser menor que un umbral.
- La intensidad de cada uno de los módulos de los puntos del borde debe ser superior a un umbral, para evitar pequeñas perturbaciones como los cambios de luminosidad, y su disparidad debe ser pequeña.
- La intensidad media de los módulos de los puntos del borde debe ser menor que otro umbral, establecido para evitar los problemas en los cambios de plano, momento en el cual los valores aumentan considerablemente.

En el caso de que no se cumpla alguna de estas condiciones, se imponen otras dos alternativas

- La diferencia entre los dos ángulos más dispares debe ser menor que un umbral.
- La media de la intensidad de los módulos de los puntos del marco debe ser superior a un umbral.

En el caso de cumplimiento de alguno de los dos grupos de condiciones, se establece que se produce movimiento de cámara, como se aprecia en la Fig. 32 (puntos exteriores en verde y puntos interiores en morado).

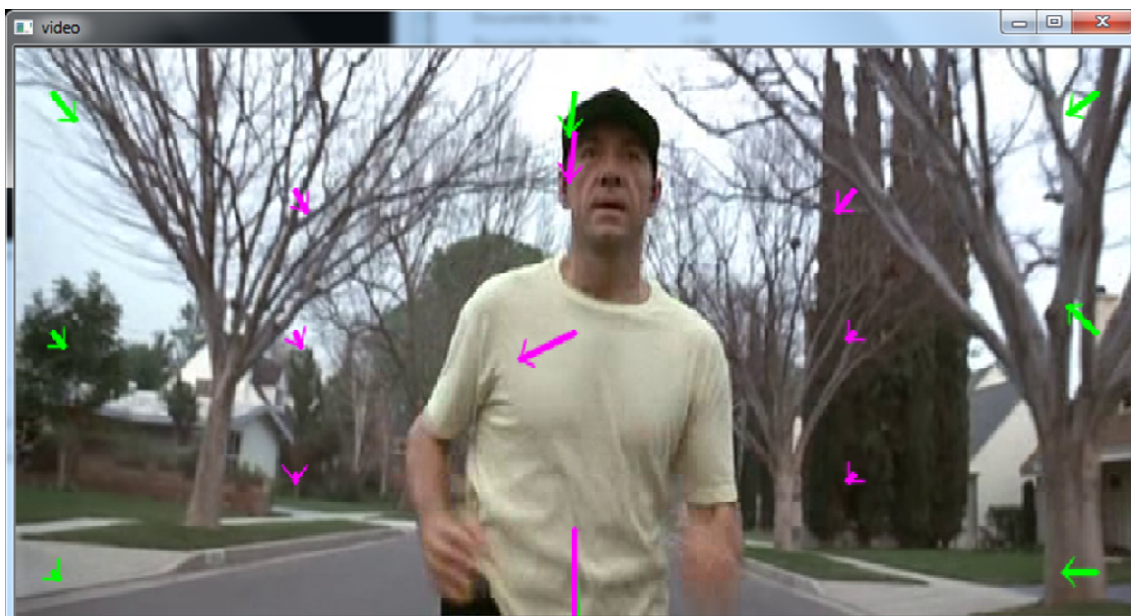


Fig. 32. Detección del movimiento de cámara y del movimiento interno de cuadro

Detección de cortes

La necesidad de extraer información sobre los vectores de movimiento relativos a todos los *frames* de un vídeo ofrece, a su vez, la posibilidad de utilizar la información de movimiento también para detectar cortes entre planos. Por ese motivo, después de comprobar que los valores de los vectores de movimiento eran muy altos en el instante en que se produce un cambio de plano mediante corte, se decidió el uso de este descriptor de bajo nivel. Esta reutilización, además, permite un aumento considerable en la eficiencia del procesamiento del vídeo, ya que otras técnicas encontradas en el estado del arte también implicaban el procesamiento de todos los *frames* del vídeo.

Es importante tener en cuenta dos factores importantes para decidir la presencia de un corte en el vídeo a partir de los valores de intensidad de los vectores de movimiento:

- La intensidad de los vectores de movimiento en un frame en el que se produce una transición por corte crece en todos los puntos de la imagen, siendo muy superior a la intensidad de los frames en continuidad.
- La intensidad no crece paulatinamente, como corresponde a los movimientos de cámara, sino que se produce de manera súbita entre un frame y el siguiente.

Este último punto resulta fundamental, ya que todas las medidas se realizan respecto a los valores anteriores de intensidad del movimiento. Esto impide la aparición de falsos positivos cuando se producen movimientos de cámara.

A partir de esta condición, las primeras pruebas se realizaron empleando simplemente el valor de la media aritmética de los puntos seleccionados para el estudio del flujo de movimiento. Estas pruebas, sin embargo, no resultaron satisfactorias, ya que en los casos en los que un objeto aparece súbitamente en la imagen a gran velocidad, la intensidad media del movimiento era equiparable a la de un corte y se producían falsos positivos.

Por esta razón, se sustituye el uso de la media aritmética por el de la media armónica, que se define a continuación.

La media armónica, denominada H , de una cantidad finita de números, es igual al recíproco o inverso, de la media aritmética de los recíprocos de dichos valores.

Así, dados n números x_1, x_2, \dots, x_n , la media armónica será igual a:

$$H = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} = \frac{n}{(\frac{1}{x_1} + \dots + \frac{1}{x_n})} \quad (79)$$

La media armónica resulta poco influida por la existencia de determinados valores mucho más grandes que el conjunto de los otros, siendo en cambio sensible a valores mucho más pequeños que el conjunto. La media armónica no está definida en el caso de que exista algún valor nulo, por lo que en ese caso se omite la introducción de este elemento.

La ventaja es que considera todos los valores de la distribución y en ciertos casos, es más representativa que la media aritmética.

Esta medida es mucho más fiable para la detección de cambios de plano por corte, y ya garantiza un nivel de acierto muy alto. Sin embargo, la frecuencia del uso de este recurso en los contenidos audiovisuales, implica que sea necesaria una precisión mayor, incluso para los casos anómalos.

Según las pruebas realizadas, los únicos problemas surgidos ante casos anómalos era la aparición de falsos positivos, por lo que se impuso una condición extra a todos los cortes considerados candidatos, que es la comprobación de la correlación entre las imágenes previa y posterior al corte. En el caso de que el corte sea real, el valor de esta correlación es muy baja. En el caso de que el falso positivo se deba a algún otro tipo de transición, como un encadenado, o al movimiento anómalo de objetos o personas dentro del cuadro, el valor de correlación resulta muy alto.

El cálculo de la correlación se ha obtenido mediante la ecuación, definida en [54].

$$c = \frac{\sigma_{xy}}{\sqrt{\sigma_{xx} \cdot \sigma_{yy}}} = \frac{\sum_{i=1}^n \sum_{j=1}^n (x_{ij} - \bar{x}) \cdot (y_{ij} - \bar{y})}{\sqrt{(\sum_{i=1}^n \sum_{j=1}^n (x_{ij} - \bar{x})^2) \cdot (\sum_{i=1}^n \sum_{j=1}^n (y_{ij} - \bar{y})^2)}} \quad (80)$$

Donde x_{ij} es el valor digital del píxel (i,j) de la primera imagen y y_{ij} es el valor digital del píxel (i,j) de la segunda imagen, siendo n el tamaño de la ventana sobre la que se buscan las similitudes de las imágenes.

Detección de fundidos

La detección de fundidos se presenta mucho más sencilla, ya que basta con comprobar que una imagen negra se funde paulatinamente con la imagen del vídeo, ya sea en un sentido o en el inverso.

Para detectar un fundido se comprueba que el número de bordes (integrado en una ventana de N *frames*, que se considera de 5 *frames* según las pruebas realizadas con fundidos reales, ya que este valor depende de la velocidad del fundido) tiene una dinámica estrictamente descendente hasta acabar en 0 bordes. Para esto se emplea el descriptor de bajo nivel de gradiente ya definido y extraído de todos los *frames*. Si la imagen en ese instante es negra consideramos que ha habido un *fade out*. Si, por el contrario, aparece una imagen en negro con 0 bordes y paulatinamente va ascendiendo el número de bordes, se detecta un *fade in*.

Detección de encadenados

El caso de la detección de encadenados es complejo, y no se ha alcanzado un resultado óptimo. Sin embargo, el uso de encadenados no es tan frecuente como para resultar determinante, y el descriptor sí permite tener una idea cualitativa del uso de este recurso en un vídeo.

Para la detección de encadenados se utiliza el descriptor del gradiente, que permite comprobar el número de bordes. Cuando se produce el encadenado, se superponen dos imágenes, con lo cual el número de bordes de la imagen superpuesta debería ser igual a la suma de los bordes de la imagen anterior y la imagen exterior. Sin embargo, la disolución de una imagen en otra implica un cambio en la nitidez de la imagen, que a su

vez condiciona el número de bordes detectado. Por lo tanto, después de una exhaustiva batería de pruebas, se consideró que la mejor opción era localizar máximos o mínimos locales en los valores de gradiente del vídeo. De esta manera, el porcentaje de acierto del descriptor es suficientemente alto para obtener la idea cualitativa deseada para la recomendación.

También se comprobó que algunas mejoras en la eficacia de la detección encarecían enormemente el coste computacional de la operación y ralentizaba el procesado del vídeo, por lo que se decidió no incorporarlas a la versión final.

Selección de *key-frames*

Para la selección de los *key-frames* se busca la imagen más representativa de un plano, por lo que el algoritmo se ejecuta cada vez que se localiza un cambio de plano, ya sea por corte, por fundido o por encadenado.

Para ello, se realiza un proceso iterativo que, sin embargo, es bastante rápido. Se escogen de forma pseudo-aleatoria (imponiendo unas restricciones de distancia temporal mínima para no desvirtuar el proceso) dos *frames* contenidos en el plano y se obtienen sus histogramas de color. Si la distancia es inferior a un umbral previamente fijado, se considera que se trata de un plano fijo, por lo que se escoge cualquiera de ellos. En caso contrario, se escoge un tercero siguiendo el mismo criterio pseudo-aleatorio con las restricciones temporales. Con los tres histogramas de color, si está dentro de un umbral mayor que el anterior, se escoge el que esté a menor distancia del histograma mediana de color; de lo contrario, se continúa el proceso iterativo.

Detección de escenas

La detección de escenas es otro proceso que puede requerir una gran carga computacional sin que los resultados aumenten de forma espectacular, por lo que se ha decidido, de forma análoga a los encadenados, utilizar este descriptor como una orientación cualitativa de las escenas que componen un vídeo.

Es importante destacar que, mientras el resto de metadatos que definen el montaje están vinculados a sus funciones sintáctica o rítmica, este es el único descriptor que encaja dentro de la función narrativa del montaje, cuya importancia es muy grande de cara al modelado de los gustos de los usuarios. Sin embargo, el avance en el procesado automático de la narrativa audiovisual requiere del uso de herramientas y recursos semánticos, lo que conforma una de las más interesantes líneas de futuro de este campo de conocimiento.

Las escenas, como ya se ha señalado, son la unidad audiovisual situada entre el plano (de significado puramente sintáctico) y la secuencia (de significado puramente semántico). La escena, por su parte, señala una unidad de acción relacionada con un determinado escenario. Por ejemplo, una secuencia de un atraco a un banco se puede dividir en la escena en la que los ladrones se juntan para preparar el plan, la escena en la que se desplazar al lugar del crimen, la escena del atraco y la escena de la huída. Cada una de estas escenas tiene un determinado sentido semántico pero, a su vez, están delimitadas por su relación con escenarios distintos. Por este motivo, una segmentación adecuada de escenas debería tener en cuenta aspectos estructurales y morfológicos de la

imagen, que permitan identificar un determinado escenario, pero también aspectos semánticos. Por lo tanto, la aproximación realizada, a partir de la identificación de escenarios, sin ser totalmente fiable, da una idea cualitativa de la acción narrativa que tiene lugar, debido a la relación habitual en el audiovisual (no intrínseca pero sí mayoritaria en el lenguaje audiovisual), de corresponder segmentos de acción con el cambio de escenarios.

Para la segmentación de escenas se utiliza un algoritmo similar al de la elección del *key-frame*, en el que los objetos de partida son los propios *key-frames* elegidos, y en el que se tienen en cuenta patrones habituales del lenguaje audiovisual. Una mera comparación de histogramas y posterior umbralización entre los *key-frames* no servía ni siquiera como primera aproximación, ya que es necesario considerar la presencia de insertos, de imágenes plano-contraplano, o de narraciones en contrapunto. Por ello, se emplea el algoritmo definido por Lin y Zhang [65], que detecta escenas mediante agrupación de planos. Para esto, se utiliza una ventana de tamaño variable, en cada una de cuyas posiciones se guardan los valores del histograma de color, y se buscan similitudes entre *frames* distantes en un número variable de posiciones. De este modo, se obtiene un ratio relativo al *frame* actual sobre un grupo de *frames* pasados y un grupo de *frames* futuros, con lo que el valor de este ratio determina la presencia o ausencia de un cambio de escena, encontrándose así los límites que agrupan los planos que conforman cada escena.

5.5 Pruebas y validación

5.5.1 Validación de descriptores de imagen

Marco de pruebas

Las pruebas del sistema se plantean como una forma de validación de los algoritmos desarrollados para el recomendador de imágenes basado en contenido. Los algoritmos de filtrado colaborativo quedan fuera de esta validación, debido a que su implementación se ha restringido al alcance actual del estado del arte, por lo que las pruebas ya realizadas son muy numerosas y suficientemente conocidas [66].

Dentro de este propósito de validación del sistema, se distinguen dos tipos de tests: los que van dirigidos a los procesos de clasificación de las imágenes, y los que tienen que ver con los procesos de recomendación. Este capítulo se centra en la validación de los procesos de clasificación: en primer lugar referida a las imágenes, en segundo lugar a los vídeos.

Para realizar esta evaluación de los descriptores para la clasificación de imágenes, se han seleccionado 200 imágenes de la galería de arte visual Ciudad de la pintura [67]. Estas imágenes corresponden con cuadros comprendidos entre los siglos XV y XX, y representan diferentes estilos artísticos, seleccionadas de ese modo con el fin de que la evaluación sea lo más completa y directa posible. Estas imágenes han sido puntuadas entre el 1 (si la imagen no gusta en absoluto) y el 5 (si el gusto por la imagen es máximo) por 60 usuarios que se han prestado como voluntarios.

Resultados del test de clasificación

El test tiene como objetivo comprobar que los descriptores definidos para la clasificación de las imágenes sirven, efectivamente, para caracterizar los gustos de los usuarios de la mejor manera posible. Además, se establece una lista de la importancia global que cada uno de los atributos tiene para el conjunto de usuarios.

Para cada usuario, se han elegido las $200 \cdot thiu$ imágenes favoritas del corpus, donde $thiu$ es el umbral que define el porcentaje de imágenes favoritas que se considera. Esta selección de imágenes favoritas se ha realizado siguiendo un procedimiento descrito dentro del algoritmo de recomendación basado en perfil, en el capítulo 6 de esta tesis. El valor seleccionado para el test ha sido $thiu=0,15$, por lo que finalmente se han elegido las 30 imágenes favoritas de cada usuario.

Partiendo de estos datos, se ha definido una medida denominada “Importancia”.

Siendo a cualquiera de los descriptores de clasificación definidos para el sistema, se ha definido la medida de la “Importancia” comparando el valor de la varianza del atributo a dentro del grupo de imágenes que gustan a cada usuario con la varianza de ese mismo atributo dentro del conjunto de todas las imágenes:

$$\text{Importancia}_a = \sum_{\text{user}=1}^{60} \frac{\text{Var}(X_a)}{\text{Var}(Y_{au})} \quad (81)$$

Donde X corresponde a los valores del descriptor en las 200 imágenes del corpus e Y a los valores del descriptor en las imágenes favoritas del usuario u .

También se ha definido otro parámetro adicional, “Usuarios afectados”, que establece cuántos usuarios de los 60 iniciales tienen sus gustos influenciados por ese descriptor. La manera de obtener ese dato consiste en contar el número de usuarios que cumple la condición (82):

$$\frac{\text{Var}(X_a)}{\text{Var}(Y_a)} > \text{umbral} \quad (82)$$

Para los tests de la Tabla 19, se ha considerado un umbral de valor 1,15. La tabla muestra los descriptores ordenados según su importancia global. También se ha establecido un ratio como cociente de los resultados de “Importancia” y “Usuarios afectados”.

<i>Tipo</i>	<i>Descriptor</i>	<i>Importancia</i>	<i>Usuarios afectados</i>	<i>Ratio (Im/UA)</i>
Textura	Variedad de entropía	213,79	22 (37%)	9,72
Textura	Homogeneidad de líneas	160,2951	44 (73%)	3,64
Color	Variedad cromática	138,0416	43 (72%)	3,21
Luminancia	Entropía por planos de bits (4)	100,0674	56 (93%)	1,79
Textura	Energía de líneas	91,4375	45 (75%)	2,03
Color	Planaridad de color	74,1061	40 (67%)	1,85
Color	Saturación HSV (varianza)	73,4121	36 (60%)	2,04
Luminancia	Entropía por planos de bits (1)	68,5153	41 (68%)	1,67
Color	Saturación HSV (media)	63,7606	44 (73%)	1,45
Luminancia	Luminancia directa (varianza)	60,5754	37 (62%)	1,64
Color	Variedad cromática (inter)	44,1947	31 (52%)	1,43
Color	Coherencia espacial	41,7862	26 (43%)	1,61
Luminancia	Entropía por planos de bits (3)	27,8353	21 (35%)	1,33
Color	Variedad cromática (N)	20,6617	15 (25%)	1,38
Luminancia	Entropía por planos de bits (2)	17,2227	12 (20%)	1,44
Luminancia	Luminancia directa	5,3371	4 (7%)	1,33

Tabla 19. Test de validación de descriptores para la clasificación

5.5.2 Validación de descriptores de vídeo

Creación del entorno de pruebas

En el caso de los vídeos se cuenta con bases de datos masivas utilizadas internacionalmente para las pruebas y validaciones de sistemas de recomendación. Aunque existen varias opciones disponibles, se ha elegido la última versión de la más popular, que utiliza la mayor parte de los investigadores en sistemas de recomendación: se trata de la base de datos de *MovieLens*, recabada durante varios años por el grupo de investigación *GroupLens*, de la Universidad de Minnesota [68] a partir de la opinión de usuarios reales sobre una amplia base de datos de películas.

Concretamente, se ha utilizado la base de datos *MovieLens 10M*, que contiene 10681 películas, 71567 usuarios y 10000054 de valoraciones de los usuarios sobre las películas. Es importante insistir en que se trata de puntuaciones explícitas de usuarios reales sobre las películas que han visto y sobre las que, por lo tanto, pueden expresar su preferencia. Un sencillo cálculo arroja la cifra de que, en promedio, cada usuario de la base de datos ha valorado unas 139 películas. Esto podría ser un problema para la prueba de los algoritmos de la tesis por las razones que se comentan a continuación, pero, afortunadamente, este dato presenta una enorme varianza, y existen usuarios con muy pocas valoraciones, apenas una decena, y otros cuyos *ratings* se cuentan por miles.

Dado que en esta tesis se valida un sistema de recomendación basado en contenido, para ello es necesario, previamente, extraer de cada una de las películas del corpus los descriptores definidos, lo cual implica un procesamiento masivo de información. Por lo tanto, es necesario elegir un número acotado de las películas presentes en la base de datos de *MoviLens*, y llevar a cabo su procesamiento para obtener los descriptores que permitan comprobar la eficiencia tanto de la definición de descriptores como de los algoritmos de recomendación. Del mismo modo que en la recomendación de imágenes, se escogen 200 películas disponibles que, además, sean útiles para comprobar distintas funcionalidades del motor de recomendación.

Como se verá en el estado del arte del capítulo 6 de la tesis, el número de valoraciones de los ítems limita el funcionamiento de ciertos algoritmos de recomendación. Por lo tanto, para la selección de las películas del corpus, se crearon unas horquillas de número de valoraciones, de tal modo que en el corpus elegido contenga películas valoradas masivamente y películas con muy pocas votaciones. Así se comprobará también que el sistema sea capaz de valorar películas con escaso número de votaciones. Las horquillas se definen en la Tabla 20

Número de valoraciones	Mínimo número de películas	Número real de películas seleccionadas
>10000	10% (20 películas)	16.5% (33 películas)
5001-10000	15% (30 películas)	17.5% (35 películas)
501-5000	25% (50 películas)	30% (60 películas)
51-500	15% (30 películas)	24% (48 películas)
1-50	10% (20 películas)	12% (24 películas)

Tabla 20. Número de películas según el número de valoraciones

La lista de las 200 películas incluidas finalmente en el corpus, junto con su número de valoraciones, se puede consultar en el Anexo IV.

Una vez que las 200 películas fueron conseguidas físicamente y almacenadas en una base de datos interna creada para la realización de las pruebas, fue necesario proceder a la adaptación de la base de datos de *MovieLens* y al cruce de información con los datos extraídos de las películas seleccionadas.

Para ello, en primer lugar, se insertó la información de *MovieLens* en una base de datos MySQL instalada en un PC servidor configurado para la realización de las pruebas. A continuación, se seleccionaron, de la base de datos original de *MovieLens*, las tres tablas necesarias: *ítems* (para identificar las películas), *users* (para identificar los usuarios) y *ratings* (donde se recogen las valoraciones, cada una de las cuales viene dada para un

usuario y una película). De estas tres tablas se conservó la estructura, pero se eliminaron los datos innecesarios para ganar eficiencia. En el caso de *users*, el contenido de la tabla se podía mantener, ya que el objetivo era tener el máximo número posible de usuarios para que la validación sea más fiable; sin embargo, en el caso de *ítems*, solo era necesario mantener la identificación de las 200 películas seleccionadas para el corpus, por lo que el resto de información se eliminó. Del mismo modo, en la tabla *ratings* (que era la mayor y la que más ralentizaba las consultas a la base de datos) se eliminaron todas aquellas puntuaciones referidas a las películas presentes en la base de datos original de *MovieLens* pero que no existían en el corpus seleccionado. De este modo, la base de datos quedó finalmente establecida con 952363 ratings, de 71567 usuarios para 200 películas.

Una vez realizada la adaptación, el último paso necesario fue añadir la estructura de las tablas en las que se guardan los valores de los metadatos finales extraídos de cada película. Para ello, se incluyeron dos tablas: una de ellas, *vdcriptors*, guarda los valores absolutos de los descriptores finales, y la otra *vdcriptorsnorm*, guarda esos mismos valores normalizados a la misma escala. Esta normalización se realiza simplemente llevando los valores a la escala común, entre 0 y 1, según la fórmula (82) para facilitar las operaciones y poder realizar las comparaciones de los descriptores. Otras posibles normalizaciones necesarias tienen lugar en la fase de recomendación. El método de normalización elegido para este proceso se debe a que mantiene la proporcionalidad del rango de elementos.

$$\hat{d}_i = \frac{d_i - \min(d_i)}{\max(d_i) - \min(d_i)} \quad (83)$$

Donde \hat{d}_i es el valor normalizado del descriptor d_i . Por lo tanto, las tablas de la base de datos necesarias para comprobar la influencia de los descriptores definidos en los usuarios son 5 (*ítems*, *users*, *ratings*, *vdcriptors* y *vdcriptorsnorm*), y se representan en el diagrama de la Fig. 33.

Una vez configurada la base de datos necesaria, se procedió a la anotación de las 200 películas (cada una identificada en el nombre del archivo por el identificador correspondiente de la base de datos original de *MovieLens*), para lo cual se creó un programa manejador que iba analizando las películas disponibles, una tras otra, e insertando los valores de los descriptores en la tabla *vdcriptors*, relacionando cada entrada con el identificador de la tabla *ítem*, que se conserva respecto a la base de datos *MovieLens* y permite mantener sincronizada la información de las valoraciones.

Una vez anotadas las 200 películas, se procedió a la normalización, con lo que se rellenó la tabla *vdcriptorsnorm*. A partir de los datos de esta tabla y las valoraciones de los usuarios, se realizan las pruebas de influencia de los descriptores que se detallan en los siguientes apartados.

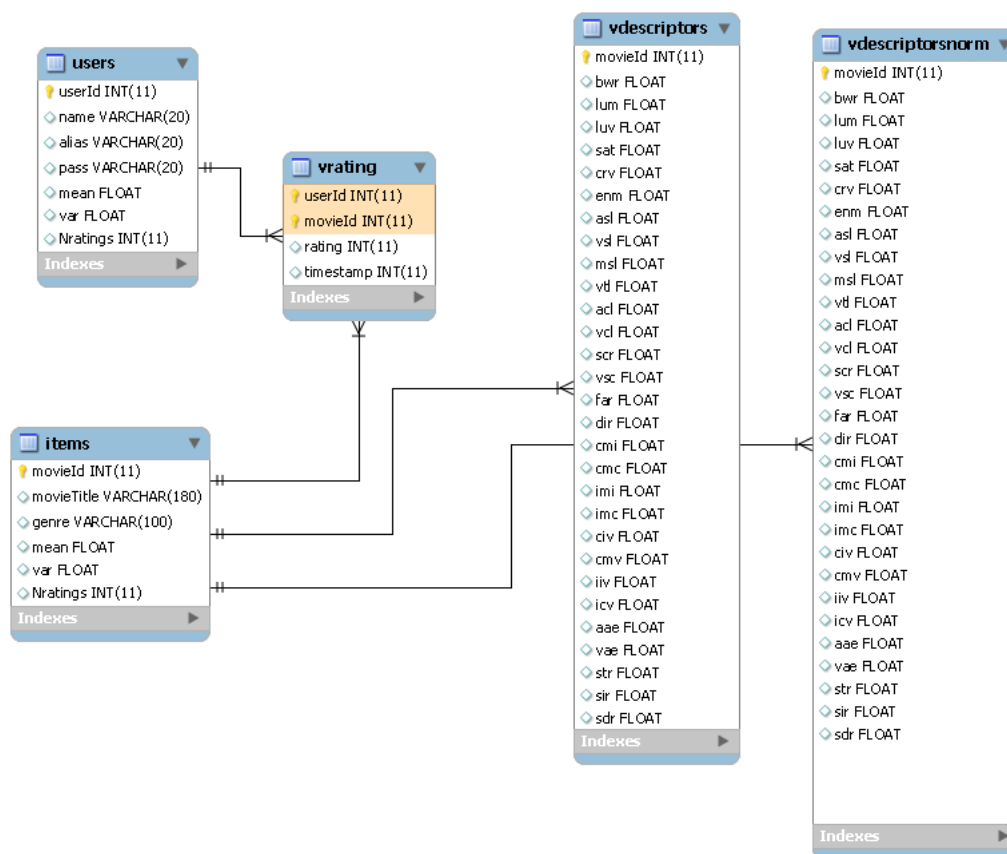


Fig. 33. Base de datos necesaria para la validación de los descriptores de vídeo

Contraste de hipótesis Kolmogorov-Smirnov

Como se aprecia en las gráficas de los histogramas de los descriptores (Fig. 34), cada uno de ellos **tiene diferentes distribuciones de probabilidad**, por lo que es necesario recurrir a herramientas más generales que en el caso de las imágenes para comparar las homogeneidades de las distribuciones de los valores de los descriptores aplicados tanto a todas las películas como a aquellas que son del gusto del usuario.

El objetivo de la prueba es comprobar, para cada usuario de la base de datos que ha valorado un mínimo número de películas entre las 200 del corpus, qué descriptores condicionan sus preferencias, para lo cual se sigue un razonamiento inverso: se comprueba qué descriptores modifican su distribución de probabilidad después de aplicarles el filtro del gusto del usuario. Hay que considerar que el usuario se comporta como un filtro multidimensional, de tal manera que las películas que le gustan son aquellas que cumplen los parámetros de ese filtro.

Dado que las distribuciones de probabilidad son diferentes para cada descriptor, se opta por aplicar la prueba de contraste de hipótesis de Kolmogorov-Smirnov, que comprueba la homogeneidad de dos distribuciones probabilísticas y determina, para un determinado nivel de significación, si las variaciones entre ambas distribuciones se deben al azar o a que, efectivamente, son muestras de dos distribuciones distintas: por un lado, la probabilidad de ocurrencia en una película cualquiera y, por otra, la probabilidad de ocurrencia en las películas del gusto del usuario.

El contraste χ^2 tiene el problema de que no considera el valor exacto de las observaciones, por lo que solo puede usarse con tamaños muestrales apreciables. Por esta razón se recurre al test de Kolmogorov-Smirnov

El test de Kolmogorov-Smirnov se formaliza de la siguiente manera para comprobar la influencia del descriptor D en el usuario U:

Se consideran dos muestras aleatorias simples; una de ellas, son los valores del descriptor D para las 200 películas del corpus: $X_1, X_2 \dots X_n$, donde $n=200$; la otra, los valores de los descriptores de las películas que gustan al usuario U: $Y_1, Y_2 \dots Y_m$. A partir de aquí, el test de Kolmogorov-Smirnov intenta dilucidar si las distribuciones poblacionales desconocidas F_X y F_Y coinciden, lo cual equivale a contrastar la hipótesis nula $H_0: F_X = F_Y$.

Considerando las distribuciones muestrales siguientes:

$$F_n^*(x) = \frac{1}{n} \sum_{i=1}^n I\{X_i \leq x\} \quad (84)$$

$$G_m^*(x) = \frac{1}{m} \sum_{i=1}^m I\{Y_i \leq x\} \quad (85)$$

Se puede formar el estadístico

$$\Delta_{n,m} = \sup_{x \in \mathbb{R}} |F_n^*(x) - G_m^*(x)| \quad (86)$$

Que mide la diferencia global existente entre ellas y se denomina estadístico de Kolmogorov-Smirnov correspondiente a ambas muestras.

Si la hipótesis H_0 es cierta, es probable que $F_n^*(x)$ y $G_m^*(x)$ sean próximas (puesto que ambas constituyen estimaciones de la distribución poblacional común) y que $\Delta_{n,m}$ tenga, por tanto, un valor relativamente pequeño. En cambio, si $F_X \neq F_Y$, cada una de las dos distribuciones muestrales tenderá a aproximarse a su distribución poblacional y el valor de $\Delta_{n,m}$ será más alto. Esto lleva a contrastar la hipótesis H_0 , mediante un test con región crítica de la forma:

$$\{\Delta_{n,m} > k\} \quad (87)$$

Y es necesario, por lo tanto, estudiar la distribución de $\Delta_{n,m}$ para poder determinar el punto crítico k asociado a cada nivel de significación.

Aplicado al problema de dilucidar si el descriptor D es influyente para el usuario U, esto ocurrirá cuando el valor del estadístico $\Delta_{n,m}$ sea superior al valor k , que para un determinado grado de significación p viene dado por la tabla de distribución de Kolmogorov-Smirnov. Dado el valor de $n=200$, ya que el corpus se compone de 200 películas anotadas, la distribución del estadístico se ajusta a las siguientes fórmulas para valores de $n > 16$ [69]:

$$P\left\{\Delta_{n,m} > 1.07 \sqrt{\frac{n+m}{n \cdot m}}\right\} = 0.2 \quad (88)$$

$$P\left\{\Delta_{n,m} > 1.22 \sqrt{\frac{n+m}{n \cdot m}}\right\} = 0.1 \quad (89)$$

$$P\left\{\Delta_{n,m} > 1.36 \sqrt{\frac{n+m}{n \cdot m}}\right\} = 0.05 \quad (90)$$

$$P\left\{\Delta_{n,m} > 1.52 \sqrt{\frac{n+m}{n \cdot m}}\right\} = 0.02 \quad (91)$$

$$P\left\{\Delta_{n,m} > 1.63 \sqrt{\frac{n+m}{n \cdot m}}\right\} = 0.01 \quad (92)$$

La aplicación del test de Kolmogorov-Smirnov en el estudio de la influencia de los descriptores en las preferencias de los usuarios es inmediata excepto por una salvedad: determinar la manera en la que se escogen los datos que representan los descriptores de las películas que gustan al usuario.

En el caso de la recomendación de imágenes, el corpus de prueba estaba formado por una serie de usuarios a los que se había hecho valorar un conjunto fijo de 200 imágenes. Es decir, la elección, la presencia de una valoración, no aportaba ningún dato cualitativo acerca de las preferencias. Por ese motivo era necesario segmentar las valoraciones positivas (aquellas con una puntuación más alta) de las negativas (aquellas con puntuación más baja).

En el caso de la recomendación de vídeos, sin embargo, un usuario solo tiene valoradas aquellas películas que previamente ha visto, lo cual ya da las primeras pistas acerca de sus preferencias. Selección y consumo, como se explica en el capítulo 4 de la tesis, implica, en determinado grado, preferencia. Por esta razón, en este caso se consideran dentro de las preferencias todas las películas valoradas por un usuario, donde esa preferencia está ponderada por la puntuación otorgado, de mayor (5 puntos) a menor (1 punto). Esta idea se mantiene en el algoritmo de recomendación, como se puede comprobar en el capítulo 6.

Por lo tanto, dados el usuario u y el descriptor D , es necesario seleccionar los valores del descriptor que conformen la muestra de datos $Y_1, Y_2 \dots Y_m$. Para ello, se escogen los valores del atributo D de aquellas películas que el usuario ha valorado y, a continuación, se replica ese descriptor tantas veces como puntos haya dado el usuario en su valoración. Es decir, se considera cada puntuación del usuario como unidades de preferencia, por lo que valorar con una puntuación de 1 cinco películas con un determinado valor del descriptor D equivaldría a valorar con un 5 una sola película con ese mismo valor del descriptor. De esta forma, se consideran tanto los consumos como las preferencias explícitas para calibrar la influencia de los descriptores (y, posteriormente, para generar el algoritmo de recomendación).

La Tabla 21 muestra el ejemplo de las valoraciones de un usuario ficticio u que hubiera puntuado 4 películas. Se pretende hacer el estudio de la influencia del descriptor

“Duración media de plano”, por lo que se especifican los valores de este descriptor para las películas valoradas.

Película	Puntuación	ASL
La roca	4	2.83
Ed Wood	1	5.34
Titanic	3	4.29
Matrix	5	3.14

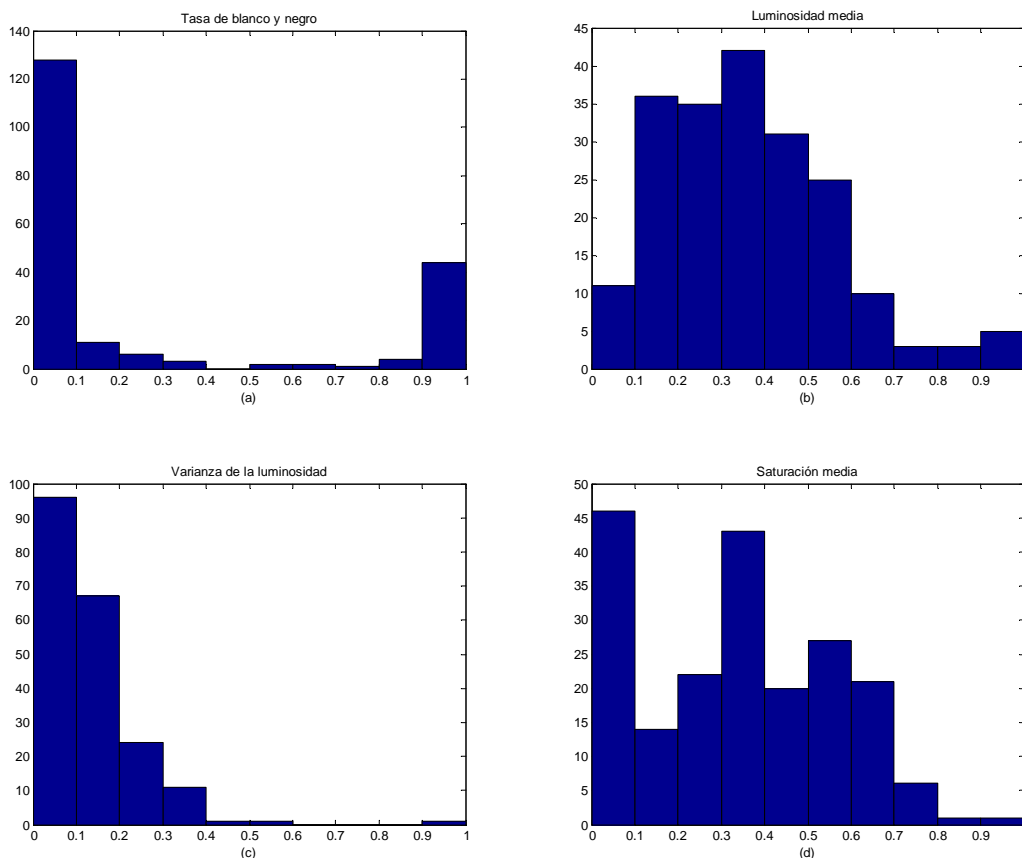
Tabla 21. Ejemplo de valoraciones del usuario u

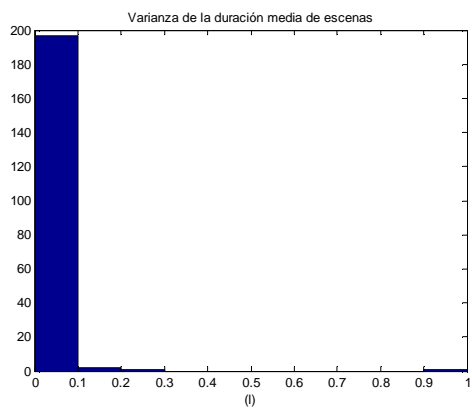
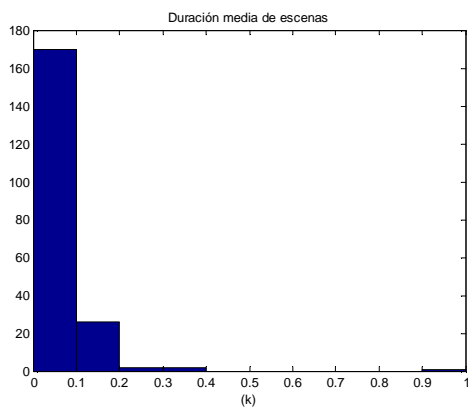
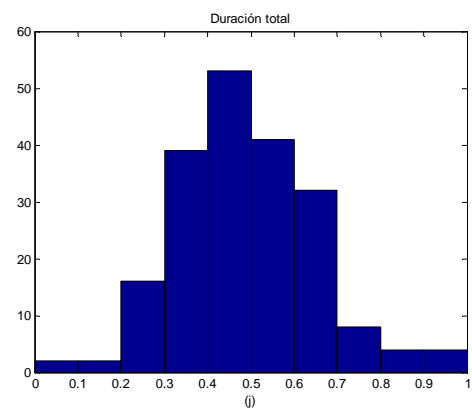
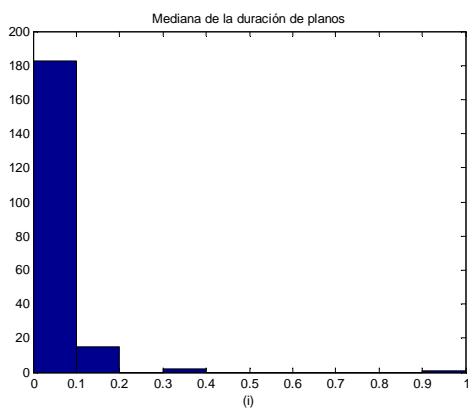
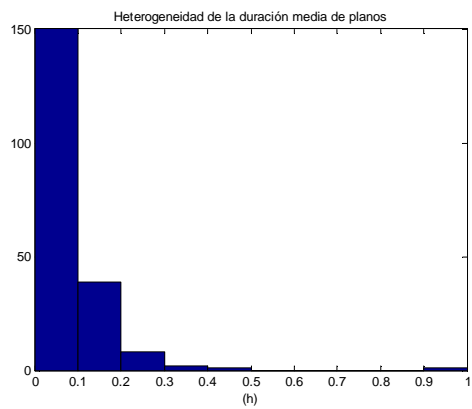
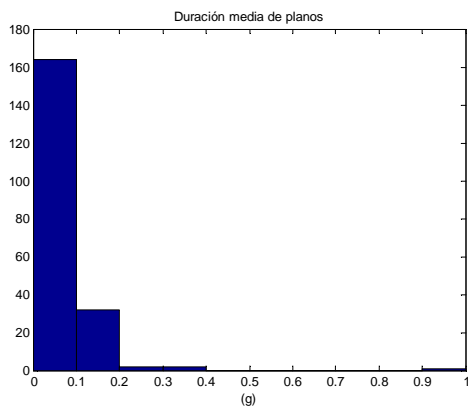
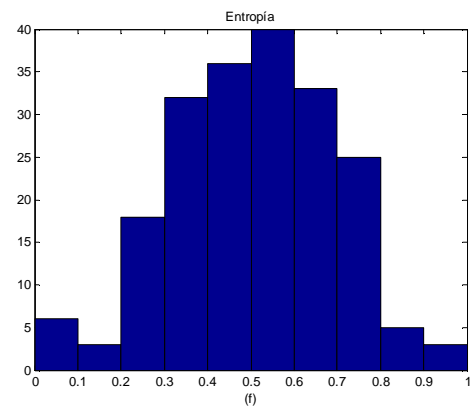
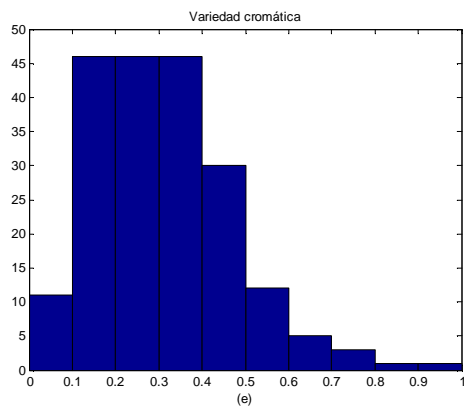
En este caso, la muestra de datos sería la siguiente: $Y_m = \{2.83 \ 2.83 \ 2.83 \ 2.83 \ 5.34 \ 4.29 \ 4.29 \ 4.29 \ 3.14 \ 3.14 \ 3.14 \ 3.14 \ 3.14\}$, donde los cuatro primeros valores corresponden a la película *La roca*, el siguiente a *Ed Wood*, los tres siguientes a *Titanic* y los cinco últimos a *Matrix*.

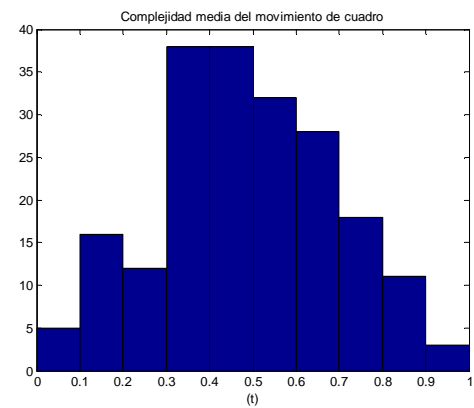
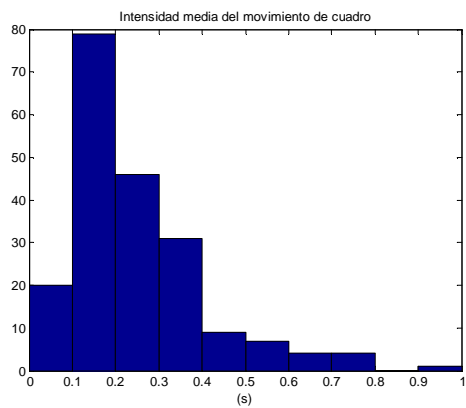
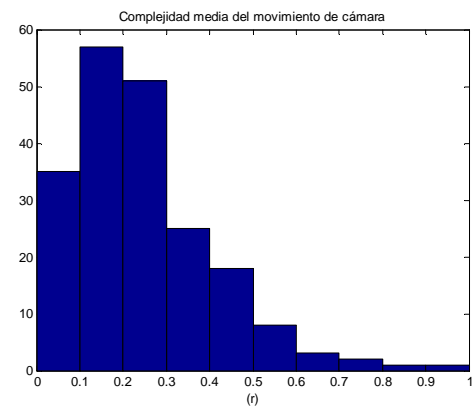
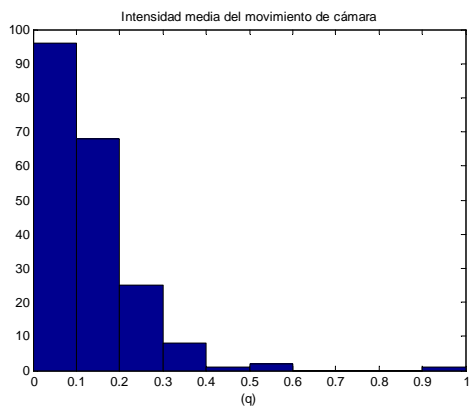
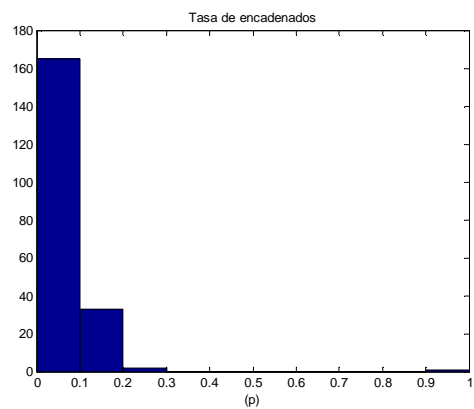
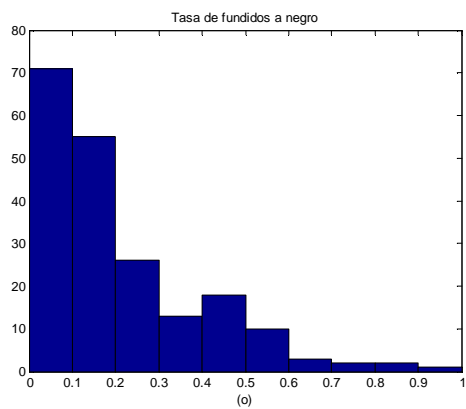
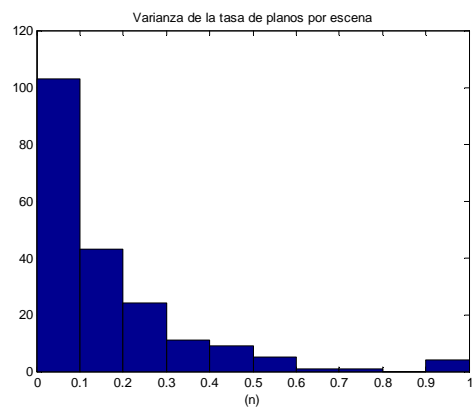
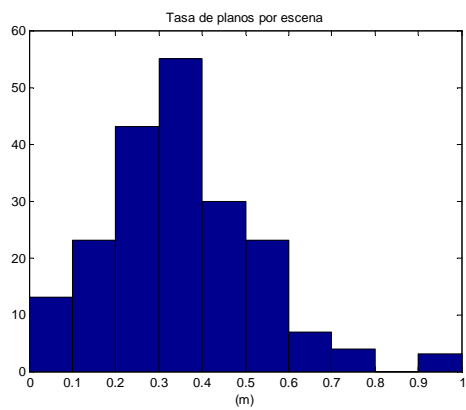
Esta muestra Y_m es la que se comparará con la muestra X_n de los 200 valores del descriptor ASL (ver fórmula (50)) para todas las películas del corpus a través del test Kolmogorov-Smirnov, que obtendrá el grado de significación a partir del cual se puede contrastar la hipótesis nula y afirmar que las dos distribuciones no se parecen, es decir, que el descriptor influye en los gustos del usuario.

Resultados

La Fig. 34 muestra los histogramas correspondientes a cada uno de los 26 descriptores definidos y, como se puede apreciar, cada uno sigue distintas distribuciones de probabilidad.







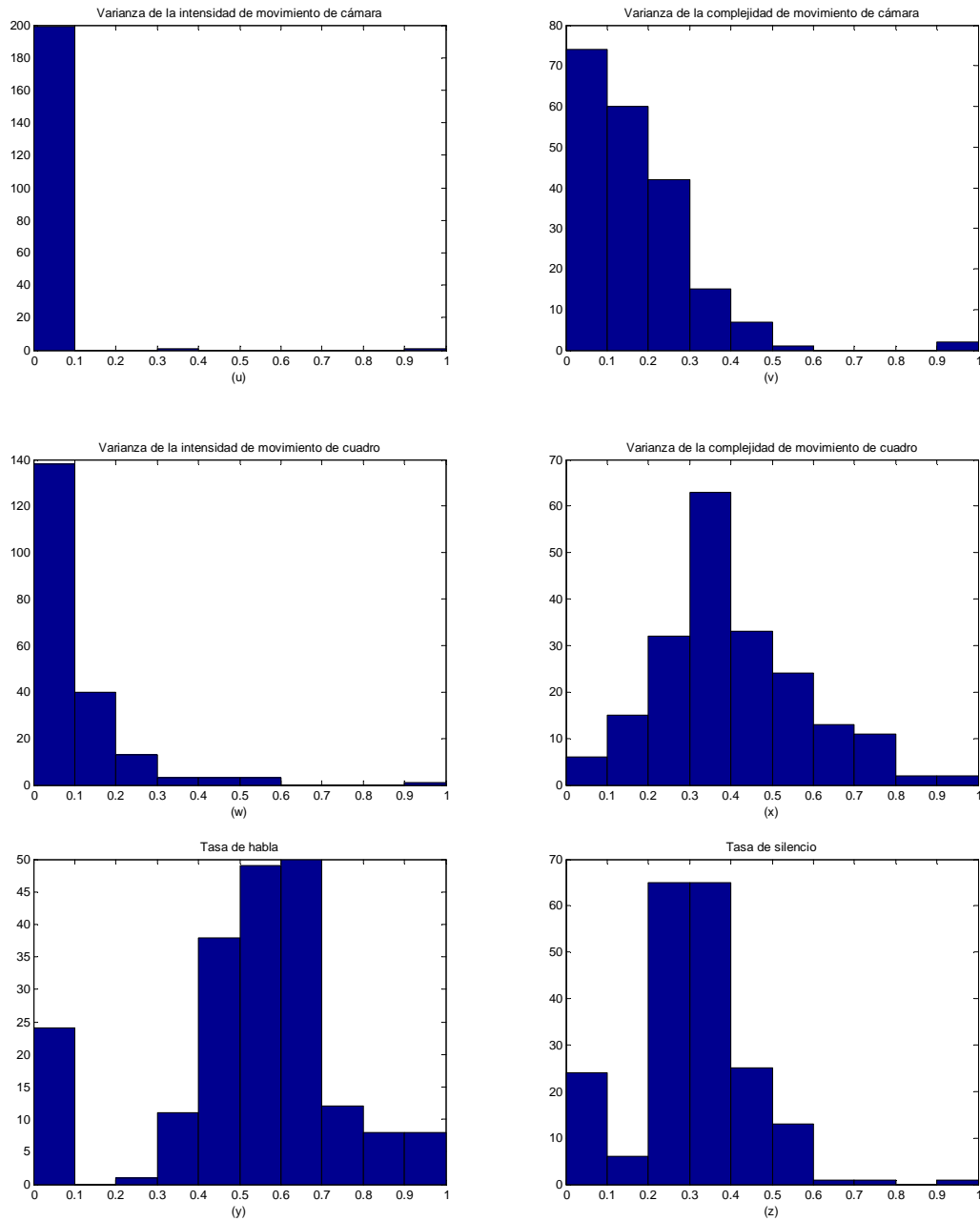


Fig. 34. Histogramas de los valores de los 26 descriptores de vídeo

Los test de Kolmogorov-Smirnov se realizaron sobre todos aquellos usuarios que tenían un mínimo de 10 películas del corpus valoradas, con el fin de evitar que la prueba quedara desvirtuada. Por lo tanto, el test final se realizó sobre 30109 usuarios.

Una vez realizados los test de Kolmogorov-Smirnov para cada uno de los 26 descriptores, promediando los valores obtenidos en cada usuario, se obtuvo la Fig. 35, que muestra el valor del estadístico, que da rápidamente una idea intuitiva de la influencia de cada descriptor para la generalidad de los usuarios.

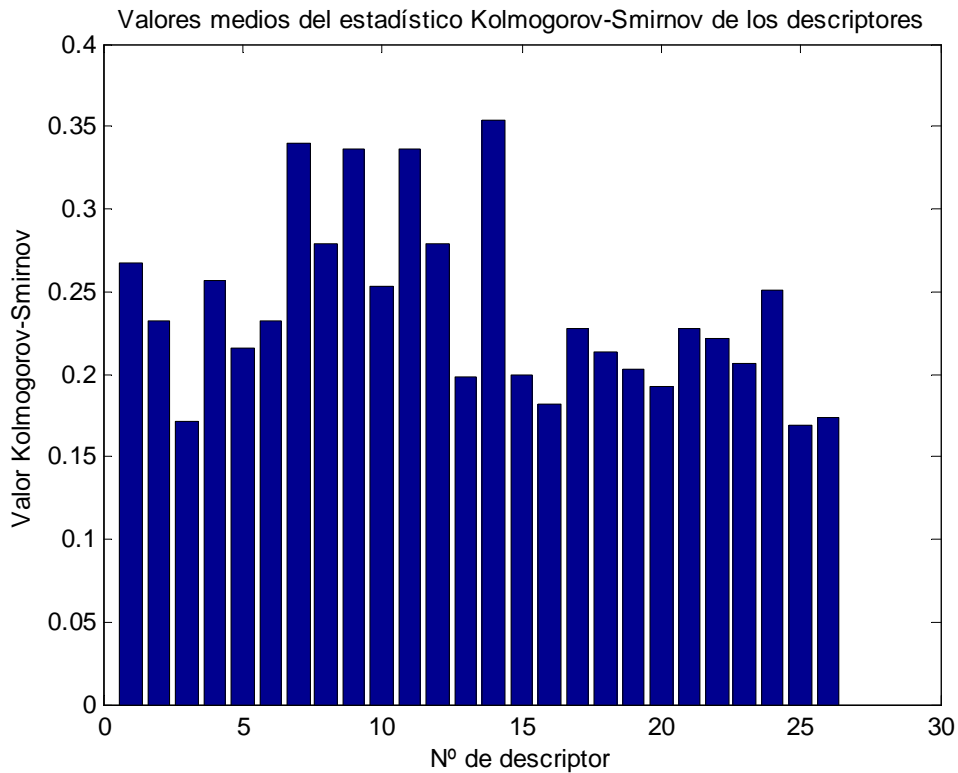


Fig. 35. Valores medios del estadístico Kolmogorov-Smirnov para los usuarios de la base de datos

Esta gráfica puede ser engañosa, porque el valor medio del estadístico no necesariamente indica a cuántos usuarios influye cada atributo, debido a que no se tiene en cuenta la varianza de los resultados. Por esa razón, se generó otra gráfica (Fig. 36) en la que se muestra el número de usuarios influido por cada uno de los descriptores para un valor de significancia de $\alpha=0.1$.

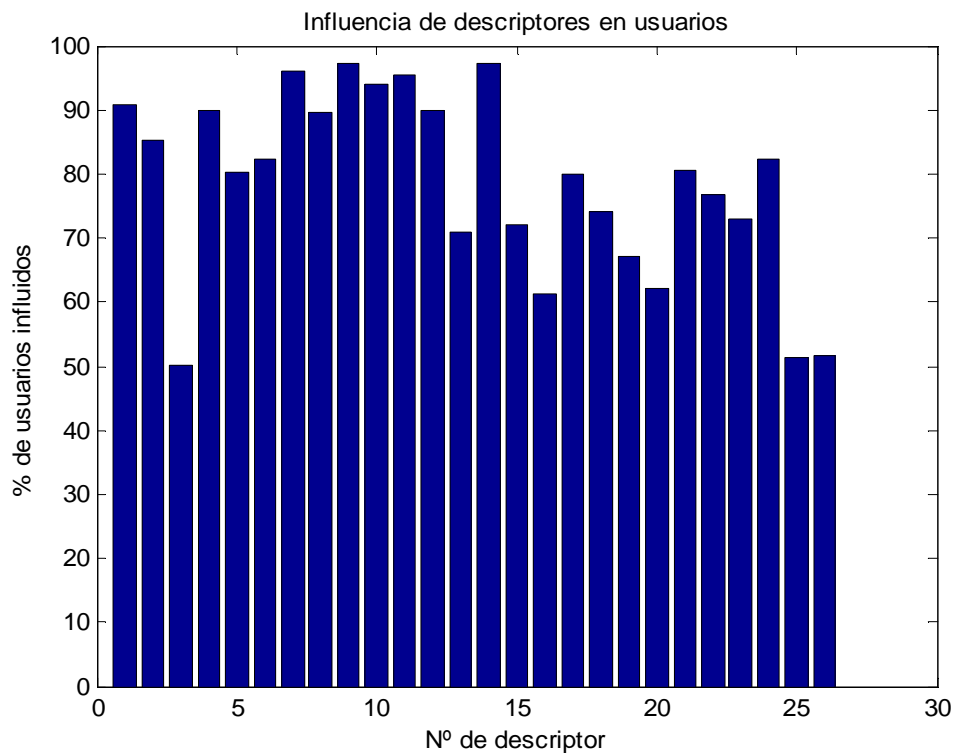


Fig. 36. Porcentaje de usuarios influidos por los descriptores del modelo

A partir de los resultados del test, cuyos resultados están recogidos en las gráficas, se valida que los 26 descriptores definidos son útiles para la recomendación. Además, se pueden extraer numerosas conclusiones a partir del análisis de las gráficas:

- Todos los descriptores definidos influyen, por lo menos, a la mitad de los usuarios.
- No se aprecian diferencias significativas entre los valores relativos del estadístico y los valores relativos del número de usuarios influidos. Es decir, no abunda el caso de que el alto valor de un estadístico se deba a una alta influencia en unos pocos usuarios, mientras para el resto de usuarios no tenga influencia. La comparación de las gráficas permite detectar este tipo de anomalías. Sin embargo, este fenómeno se aprecia con cierta intensidad en los descriptores más influyentes, observándose que el valor del estadístico saca mayor diferencia a los siguientes descriptores que la diferencia de número de usuarios a los que influyen.
- Los descriptores más influyentes están relacionados con el montaje. Por orden de importancia estos son: “Varianza de la tasa de planos por escena”, “Mediana de duración de planos”, “Duración media de escenas” y “Duración media de planos”
- El hecho de que el descriptor más influyente sea la “Varianza de la tasa de planos por escena” se puede interpretar de la siguiente manera: los usuarios son especialmente sensibles a la variabilidad del ritmo narrativo entre escenas, es decir, el hecho de que las escenas de una película se cuenten con un ritmo homogéneo o heterogéneo provoca determinadas respuestas que pueden ir desde la emoción, hasta la saturación visual o al aburrimiento. También es importante señalar que el siguiente parámetro en influencia es la duración media de planos, algo que, como se ha indicado en el estado del arte, ya había sido advertido por ciertos analistas fílmicos.
- En cuanto a los descriptores de imagen, resulta procedente destacar la importante influencia de la tasa de blanco y negro y de la saturación (lo cual es lógico, porque si el primero influye, el otro también lo hará, ya que la saturación es una propiedad del color; al contrario, sin embargo, no ocurrirá necesariamente). Esto es algo comprensible, ya que muchos usuarios, por ejemplo, descartan ver películas en blanco y negro por esa mera condición, independientemente de cualquier otro parámetro, por lo que su influencia es muy importante.
- En cuanto a los descriptores de movimiento, el más influyente es la varianza de la complejidad del movimiento interno, lo cual da una idea aproximada de la percepción del dinamismo de una escena. Cerca está también la intensidad del movimiento de cámara, que es uno de los descriptores más intuitivos y que conjuntamente con la complejidad aporta unas características estilísticas muy específicas, relacionadas con la forma y los medios de rodaje: cámara en mano, planos fijos, uso de grúas, etc.
- Por último, los descriptores menos influyentes son los relacionados con el audio, y el de la varianza de la luminosidad

5.6 Resultados obtenidos

Después de haber caracterizado las preferencias del usuario, este capítulo se ha centrado en la caracterización del contenido multimedia de cara a la recomendación. Para ello, dado que los conceptos semánticos quedan fuera del marco de la tesis, se ha recurrido a características morfológicas y estéticas de las imágenes y de los vídeos. A partir de un ejercicio de análisis de las características de la imagen y del vídeo, así como de las técnicas de medidas de calidad y de visión artificial disponibles en el estado del arte, se han definido unos descriptores que permiten caracterizar imágenes y vídeos de acuerdo con las preferencias de los usuarios. Estos descriptores, a pesar de ser en su mayoría de bajo nivel, se articulan entre ellos componiendo referencias estéticas de los objetos, ya que, como se ha demostrado en los tests estadísticos, los usuarios son sensibles a sus variaciones.

- 1) En el caso de las imágenes, la normalidad de los descriptores extraídos ha permitido comprobar su influencia en la percepción de los usuarios a través de un análisis de la varianza, demostrando que existen usuarios más sensibles que otros a ciertos parámetros de la imagen.

Para las pruebas, ante la ausencia de bases de datos estandarizadas que pudieran ser útiles para la recomendación, se recabó información de un grupo de 60 usuarios acerca de una serie prefijada de 200 imágenes.

- 2) En el caso del vídeo, los descriptores extraídos presentaban unas características probabilísticas más heterogéneas, con lo cual ha sido necesaria la utilización de una técnica de análisis distinta a la empleada con las imágenes: el test de Kolmogorov-Smirnov.

En este caso, se pudieron utilizar bases de datos públicas para la investigación en técnicas de recomendación, ya que estas utilizan películas como corpus sobre el que se recogen valoraciones de los usuarios. Por lo tanto, las pruebas de validación de la influencia de los descriptores sobre los usuarios se han llevado a cabo empleando una serie de películas recogidas en la base de datos empleada (*MovieLens 10M*).

Es importante señalar el hecho de que, tanto en el caso de las imágenes como en el de los vídeos, cada usuario estudiado se ha mostrado influido por descriptores diferentes que otros usuarios. Esto condiciona de manera determinante la idea de partida del algoritmo de recomendación, ya que es necesario tener en cuenta diferentes atributos para cada usuario, sin que estos puedan ser fijos. Esto exige que el algoritmo de recomendación realice una gestión dinámica de los criterios de recomendación basados en contenido.

6. Algoritmos de recomendación híbrida

6.1 *Introducción*

Una vez que el sistema de recomendación dispone de toda la información necesaria, relacionada tanto con los usuarios como con el contenido, es necesario aplicar los algoritmos que conforman el propio motor de recomendación. Como ya se ha explicado en los objetivos y la metodología de la tesis, los esfuerzos se han centrado en la definición de un nuevo algoritmo de recomendación basado en contenido que haga uso de los descriptores formales y estéticos definidos en el capítulo 5 y que, de esta manera, cubra el hueco del estado del arte, dado que hasta ahora los sistemas de recomendación basados en contenido han empleado únicamente palabras clave o atributos semánticos para generar las recomendaciones.

El algoritmo diseñado, por su parte, está integrado con un sistema común de extracción de metadatos de tipo estructural, capaces de describir ciertas características formales y estéticas de los contenidos audiovisuales. Sin embargo, como también se ha visto en el estado del arte, todos los tipos de técnicas de recomendación, tanto las basadas en contenido como las colaborativas, presentan ciertas deficiencias (de tipo tanto cuantitativo como cualitativo) que solo se pueden subsanar mediante la combinación de distintas técnicas de estos dos grupos. Por esta razón, el método de recomendación basado en contenido que se ha creado se aplica conjuntamente con un algoritmo de filtrado colaborativo que no supone ninguna mejora en sí mismo sobre el estado del arte (se trata de una implementación de su versión más popular y ampliamente utilizada en sistemas comerciales) pero que permite, mediante el uso de técnicas de hibridación, comprobar cuánto puede mejorar el nuevo algoritmo desarrollado, utilizándolo de forma conjunta con el filtrado colaborativo, las técnicas de uso común. Además de los resultados cuantitativos de eficiencia en las predicciones, el uso de un recomendador estético de este tipo aporta otras muchas ventajas de tipo cualitativo, que se desgranar en las conclusiones de este mismo capítulo.

Este capítulo, por lo tanto, parte de un estudio exhaustivo de los sistemas y algoritmos utilizados en el estado del arte y de una descripción del algoritmo de filtrado colaborativo empleado para hibridar con los nuevos algoritmos creados. A continuación, se describe el algoritmo aplicado tanto a imágenes como vídeos, enfocados, además, a escenarios en el que el corpus y la recogida de datos son diferentes, lo que implica incorporar ciertas modificaciones añadidas a las ya necesarias por el mero cambio del tipo de contenido, de imágenes a vídeos. La novedad investigadora de este capítulo está recogida en ese algoritmo de recomendación de imágenes y vídeos basado en contenido a partir de características formales y estéticas. Después, se describen las distintas técnicas de hibridación que se han empleado para realizar las pruebas. Finalmente, se validan los algoritmos y se realizan diversas comparativas explicando las pruebas realizadas tanto sobre el algoritmo de recomendación de imágenes como sobre el de vídeos. A modo de coda, el apartado de

conclusiones explica la manera en que los nuevos algoritmos mejoran, desde el punto de vista cuantitativo y cualitativo, los sistemas previamente estudiados en el estado del arte.

La idea central sobre la que se proyecta el nuevo algoritmo de recomendación basado en contenido es la de la variabilidad de la percepción humana entre diferentes personas y la afinidad a diferentes estilos artísticos por parte de un único individuo. Dado que estos estilos artísticos son una construcción artificial creada a partir de la similitud de ciertos patrones formales, teóricos y conceptuales, el sistema tiene como objetivo, a partir de las características formales de imágenes y vídeos, crear una propia e individualizada clasificación estética según la percepción del usuario. Considerando que el gusto estético (no así los estilos estéticos) de los usuarios no se puede definir a priori, el sistema analiza las preferencias de cada usuario, halla los patrones comunes de aquello que resulta más afín o más alejado de sus propias preferencias y, a partir de ahí, construye un sistema de coordenadas estético único para ese usuario. Por lo tanto, los objetos del corpus se representan de manera diferente para cada usuario, teniendo en cuenta las características perceptivas manifestadas por sus preferencias.

Quizás uno de los grandes problemas de los sistemas de recomendación actuales es considerar los mismos criterios para todos los usuarios. Una representación dinámica y personalizada de los elementos del corpus permite acabar con esa homogeneidad y abre enormes posibilidades de exploración de las preferencias de los usuarios. La idea central, por lo tanto, es que aunque un ítem esté, de inicio, descrito por una serie de metadatos comunes (definidos en el capítulo 5), no todos esos descriptores son percibidos de la misma manera por los usuarios. Unos usuarios, por ejemplo, serán más sensibles a las variaciones de luminosidad, otros, por el contrario, experimentarán un mayor impacto ante determinados tipos de montaje (rápidos o lentos, analíticos o sintéticos), etc. Por lo tanto, es necesario recomendar a cada usuario en función de aquellos atributos o aquella combinación de atributos que realmente afecte a su valoración estética del contenido audiovisual.

El objetivo del algoritmo, por lo tanto, es la creación de un sistema capaz de representar los elementos de forma personalizada, de tal manera que el corpus se represente en cada caso dentro del subespacio vectorial del usuario, y únicamente sobre este subespacio vectorial se articulen todas las métricas, estadísticas y otros procesos de tratamiento de datos como el *clustering*,

Este capítulo, por lo tanto, supone la culminación del proceso completo de recomendación iniciado en los capítulos anteriores, a partir del modelado de la información de usuario y de la información del contenido. La validación, por lo tanto, de los algoritmos finales, implica también la validación de algunos de los algoritmos previos. Concretamente, el modelado de los descriptores de contenidos audiovisuales explicado en el capítulo 5 supone la base de estos algoritmos de recomendación, por lo que, a pesar de que estos descriptores ya hayan sido validados en cuanto a su influencia en los gustos de los usuarios, en este capítulo se realiza, de manera implícita, la validación de su utilidad para la recomendación.

El caso del modelado del consumo audiovisual, sin embargo, es más complejo. Si bien los algoritmos de recomendación que se desarrollan en este capítulo pueden operar directamente sobre el modelo de consumo audiovisual ya explicado en el capítulo 4, los

datos de consumo obtenidos no serían suficientes para validar de forma adecuada los algoritmos de recomendación. La validación de los algoritmos de recomendación precisa el uso de bases de datos masivas de usuarios, ya que debe corroborar que su funcionalidad es aplicable a una generalidad de usuarios, por lo que se hace uso de las bases de datos públicas disponibles a tal efecto. Para poder alcanzar una cantidad de datos similar (72000 usuarios y 10 millones de *ratings*) haciendo uso del modelo de inferencia de interés a partir del consumo propuesto, haría falta realizar un despliegue comercial masivo, con miles de usuarios implicados y aparatos de medida funcionando durante largo tiempo.

No obstante, la validación del modelo de consumo llevada a cabo en el capítulo 4, junto con la validación del modelo de recomendación híbrida (que implica la validación de descriptores del capítulo 5), implica la validación total de los sistemas propuestas en esta tesis para el modelado de sistemas multimedia para personalización y recomendación híbrida a partir del consumo audiovisual de los usuarios.

6.2 Estado del arte

6.2.1 Recomendación basada en el comportamiento social

Filtrado colaborativo

El filtrado colaborativo es el algoritmo más utilizado en los sistemas de recomendación actuales, especialmente en portales de Internet, en los que resulta sencillo aglutinar información de un gran número de usuarios, y se caracteriza por dejar de lado la clasificación de los ítems (todo aquello que se puede recomendar), apostando por el juicio humano [70][71][72]. El fundamento es generar recomendaciones basadas en las opiniones que han sido expresadas por la comunidad de usuarios: “es aquel sistema en el que las recomendaciones se hacen basándose solamente en el grado de similitud entre usuarios” [73]. Por lo tanto, se trata de un filtrado de orientación cualitativa. Los filtros colaborativos se basan en la búsqueda de asociaciones entre los usuario o entre los productos generadas a partir del uso y actividad registrados por parte del usuario en un servicio. Estas asociaciones se utilizan posteriormente para realizar la recomendación en función de la afinidad o similitud que tienen los usuarios o los productos entre sí. Esta afinidad se calcula a partir de la evidencia explícita sobre las preferencias de los usuarios, como pueden ser las valoraciones concretas y adquisiciones de productos por parte los usuarios, y a partir de la evidencia implícita, que consiste, dependiendo del escenario de aplicación, en utilizar las interacciones que el usuario realice en su utilización normal de un producto para deducir el interés que suscitan determinados ítems.

La predicción de la valoración de un ítem desconocido para un usuario se basa en dos aspectos que se complementan, pudiendo partir de uno de ellos para llegar al otro y viceversa:

1. La afinidad de los usuarios: se comparan los usuarios según sus apreciaciones comunes sobre el conjunto de los ítems.
2. La afinidad del ítem: se comparan los ítems según las valoraciones que han recibido de los usuarios.

Por lo tanto, al ser el objetivo del filtrado colaborativo predecir la valoración p_{ai} del usuario a sobre un ítem i y basarse en usuarios activos, una de las desventajas radicaría en el denominado *Cold start* [74], comentado posteriormente, según el cual el sistema carece de datos iniciales para producir nuevas recomendaciones.

De acuerdo con Breese [75] podemos dividir este grupo de recomendadores en dos categorías: los métodos basados en memoria o heurísticos y los basados en modelos.

a) Filtrado colaborativo basado en memoria

El filtrado colaborativo basado en memoria divide su proceso de recomendación en tres etapas.

En primer lugar, es necesario formar un vecindario, es decir, encontrar un subconjunto único de la comunidad para cada usuario, identificando otros con intereses similares que actúen como recomendadores.

A continuación hay que agregar opiniones. Esto consiste en obtener un conjunto de pesos de los recomendadores para predecir la valoración de los ítems que aún no han sido calificados. Como es lógico, los vecinos con un mayor índice de afinidad tendrán más peso en la predicción.

Finalmente, se generan y ordenan las predicciones de los ítems, que se muestran al usuario por orden de afinidad realimentando y reiniciando el ciclo.

➤ Acercamiento basado en usuario

En primer lugar hay que definir una medida de similitud, que va a cuantificar el grado de afinidad entre dos usuarios. Una vez obtenida esta medida, sólo hay que ordenar los resultados para obtener un vecindario fiable, con los usuarios más cercanos en gustos. A partir de ahí, se puede realizar la predicción del ítem desconocido a partir del sumatorio de valores con rating [76]:

$$p_{ai} = \frac{\sum_{u \in T_a | i \in S_u} \text{sim}(a, u) \cdot r_{ui}}{\sum_{u \in T_a | i \in S_u} \text{sim}(a, u)} \quad (93)$$

Donde:

p_{ai} es la predicción de la valoración del usuario a sobre el ítem i

$\text{sim}(a, u)$ es la similitud entre los usuarios a y u

r_{ui} es la valoración del usuario u sobre el ítem i

T_a es el conjunto de vecinos de un usuario a

S_u es el subconjunto de ítems que el usuario u ha calificado

Otra posible fórmula para la predicción consiste en realizar un sumatorio de las desviaciones de la media, lo que se considera más efectivo [77], ya que compensa parte de la subjetividad de los rangos de las valoraciones:

$$p_{ai} = \bar{r}_a + \frac{\sum_{u \in T_a | i \in S_u} \text{sim}(a, u) \cdot (r_{ui} - \bar{r}_u)}{\sum_{u \in T_a | i \in S_u} \text{sim}(a, u)} \quad (94)$$

Donde:

$$\bar{r}_u = \frac{\sum_{i \in S_u} r_{ui}}{|S_u|} \quad (95)$$

➤ **Acercamiento basado en ítem**

En este caso se predice la valoración para un ítem utilizando las valoraciones del usuario en los vecinos del ítem seleccionado. Para ello maneja un vector que contiene las valoraciones de los usuarios, por lo que es necesario definir *clusters* de vecindarios para los ítems.

Análogamente al acercamiento basado en usuario, existe una fórmula basada en el sumatorio de las valoraciones:

$$p_{ai} = \frac{\sum_{j \in S_a \cap T_i} \text{sim}(i,j) \cdot r_{aj}}{\sum_{j \in S_a \cap T_i} \text{sim}(i,j)} \quad (96)$$

Donde:

p_{ai} es la predicción de la valoración del usuario a sobre el ítem i

$\text{sim}(i, j)$ es la similitud entre los ítems i y j

r_{aj} es la valoración del usuario a sobre el ítem j

T_i es el conjunto de vecinos de un ítem i

S_a es el subconjunto de ítems que el usuario a ha calificado

Y otra basada en el sumatorio de las desviaciones de la media:

$$p_{ai} = \bar{r}_i + \frac{\sum_{j \in S_a \cap T_i} \text{sim}(i,j) \cdot (r_{aj} - \bar{r}_j)}{\sum_{j \in S_a \cap T_i} \text{sim}(i,j)} \quad (97)$$

$$\bar{r}_i = \frac{\sum_{u \in U | i \in S_u} r_{ui}}{|\{u \in U | i \in S_u\}|} \quad (98)$$

- **Medida de la similitud**

Un problema fundamental de los algoritmos anteriores estriba en el cálculo de la similitud, ya sea entre dos usuarios o entre dos ítems. Existen diversas aproximaciones para obtener este valor, pero las dos más populares son las basadas en la correlación o en el coseno.

En primer lugar, se define S_{xy} como el conjunto de los ítems valorados por los usuarios x e y . Este conjunto se utiliza principalmente en sistemas de recomendación colaborativos como un resultado intermedio para calcular los vecinos más cercanos del usuario x , y se obtiene directamente, por ejemplo, mediante la intersección de los conjuntos S_x y S_y . Sin embargo, algunos métodos, como la aproximación gráfico-teórica a los filtrados colaborativos [78], pueden determinar los vecinos más cercanos sin calcular S_{xy} para todos los usuarios y .

En la aproximación basada en correlación, se utiliza el coeficiente de correlación de Pearson para medir la similitud [79][80]:

$$sim(x, y) = \frac{\sum_{s \in S_{xy}} (r_{x,s} - \bar{r}_x)(r_{y,s} - \bar{r}_y)}{\sqrt{\sum_{s \in S_{xy}} (r_{x,s} - \bar{r}_x)^2 \sum_{s \in S_{xy}} (r_{y,s} - \bar{r}_y)^2}} \quad (99)$$

En la aproximación basada en el coseno [75][81], se tratan los dos usuarios, x e y , como dos vectores del espacio m -dimensional, donde $m=|S_{xy}|$. Por lo tanto, se puede medir la similitud entre dos vectores calculando el coseno del ángulo existente entre ellos:

$$sim(x, y) = \cos(\bar{x}, \bar{y}) = \frac{\bar{x} \cdot \bar{y}}{\|\bar{x}\|_2 \times \|\bar{y}\|_2} = \frac{\sum_{s \in S_{xy}} r_{x,s} r_{y,s}}{\sqrt{\sum_{s \in S_{xy}} r_{x,s}^2} \sqrt{\sum_{s \in S_{xy}} r_{y,s}^2}} \quad (100)$$

Otra aproximación a la medida de similitud entre usuarios utiliza la diferencia media de cuadrados, y se describe en [80]. Resulta destacable que los diferentes sistemas de recomendación pueden generar diferentes aproximaciones para implementar los cálculos de similitud y valoración de la manera más eficiente posible.

b) Filtrado colaborativo basado en modelo

Este tipo de filtros colaborativos utiliza las puntuaciones realizadas por los usuarios para el aprendizaje de un modelo probabilístico, a partir del cual se pueden realizar las recomendaciones. Básicamente la idea se resume en el cálculo del *rating* $p_{u,i}$ como la probabilidad de que el usuario u dé una determinada valoración al ítem i teniendo en cuenta las valoraciones que el usuario hizo sobre otros productos.

De este modo, los filtros basados en modelo están fuertemente inspirados en las técnicas de aprendizaje automático (*machine learning*), ampliamente estudiadas en el campo de la inteligencia artificial. Los algoritmos de aprendizaje construyen el modelo probabilístico a partir del reconocimiento de patrones de actividad de los usuarios, como las puntuaciones o adquisiciones realizadas anteriormente por los usuarios. Este modelo se refina continuamente según los usuarios participen en el sistema y vayan asignando nuevas puntuaciones a los productos. A partir de las probabilidades calculadas, el sistema de recomendación es capaz de calcular el *rating* más probable de un usuario u para un ítem i .

Para la realización de este tipo de filtros se utilizan diversas técnicas. Las más empleadas en el campo de la recomendación para la construcción de los modelos probabilísticos son las siguientes:

- Redes bayesianas [82]. Esta técnica interpreta cada producto como un nodo en una red bayesiana, donde el estado de cada nodo se corresponde con los posibles valores de *rating* para cada producto. Tanto la estructura de la red bayesiana como las probabilidades condicionales se construyen automáticamente a partir de los datos.
- Creación de *clusters* de usuarios [75]. En este caso, los usuarios con la misma opinión son agrupados en clases o conjuntos de personas. Dada la pertenencia de un individuo a una clase, se puede predecir su posible valoración para un determinado producto. Tanto los parámetros como las clases del modelo se aprenden directamente a partir de los datos disponibles.

- Otras técnicas que se han empezado a utilizar recientemente son las cadenas de Markov [83] o los *Latent semantic models* [84].

La diferencia fundamental entre los filtros colaborativos basados en memoria o basados en modelos es que el primero calcula la utilidad utilizando heurísticos que aproximan la similitud entre individuos. En el segundo caso, los filtros predicen las valoraciones de los productos mediante un modelo de probabilidad construido a partir de los datos subyacentes a la actividad de los usuarios, utilizando básicamente técnicas estadísticas o de aprendizaje automático.

Problemas de los filtrados colaborativos

El filtrado colaborativo tiene, fundamentalmente, tres problemas típicos [85]:

- ***Cold start***

También es conocido como el problema del primer día [85], dado que tiene lugar el primer día de servicio, ante la falta de datos iniciales para empezar a realizar las recomendaciones. Se trata del problema más grave de los sistemas de filtrado colaborativo, y afecta tanto a los usuarios como a los ítems. Es decir, por un lado, un usuario nuevo no recibirá recomendaciones fiables hasta que el sistema conozca un número considerable de valoraciones de ese usuario; por otro lado, también existe el problema de que los nuevos ítems no pueden ser recomendados hasta el momento en que los primeros usuarios aporten sus valoraciones. Esto puede provocar que un nuevo ítem nunca llegue a ser conocido, ya que por esta razón no recibe valoraciones, y nunca será recomendado. Hasta que crece considerablemente el número de usuarios que lo valoran, el número de comparaciones multidimensionales ve seriamente afectada su escalabilidad. Por este motivo, se utilizan habitualmente técnicas demográficas para los procesos de inicialización.

- ***Banana Problem***

Los sistemas de recomendación colaborativos tienden a ser especialmente sensibles a las diferentes frecuencias de valoración de ítems por parte de los usuarios [86]. Por ejemplo, dado que las bananas son productos habitualmente adquiridos por los clientes en muchas fruterías, un sistema de recomendación que use datos de la cesta del mercado recomendará con mucha facilidad bananas. Esto no ocurrirá con un producto adquirido sólo una vez en un amplio lapso de tiempo, como por ejemplo los automóviles. En cuanto a contenidos audiovisuales, por ejemplo, saldrían beneficiados aquellos programas más populares, con más aceptación general. Por lo tanto, un usuario con gustos atípicos tendrá que conformarse con recomendaciones más pobres que las de aquello que tienen unas preferencias más usuales.

- **Efusividad y subjetividad en las valoraciones**

Las valoraciones de usuarios similares se combinan habitualmente para hacer recomendaciones de filtrado colaborativo, pero los usuarios, habitualmente, emplean diferentes escalas de valoración para expresar preferencias similares. Un usuario puede ser más efusivo en su valoración que otro y dar una mayor puntuación cuando las valoraciones deberían ser del mismo orden. Por este motivo resulta más evidente, a la

hora de efectuar las predicciones, emplear el sumatorio de las desviaciones de la media antes que los valores absolutos almacenados. A menudo se utilizan técnicas que conservan secuencias relativas y ordinales para identificar las variaciones en los patrones de valoración [87][88].

Métodos demográficos

Uno de los problemas habituales con los sistemas de recomendación basados en filtrado colaborativo es que el número de *ratings* disponible para calcular la predicción de los ítems es usualmente más pequeño en comparación con el número de *ratings* que es necesario predecir. Es decir, tenemos que recomendar para un volumen de gente mayor del que tenemos información disponible. Esto es realmente importante ya que los filtros colaborativos dependen de una masa crítica de usuarios para poder ser realmente efectivos. Por tanto, es necesario también poder realizar predicciones de *ratings* teniendo en cuenta un conjunto limitado de ejemplos e información.

Una de las soluciones más comunes para solucionar esta limitación de los filtros colaborativos (al menos, hasta que el sistema no disponga de la necesaria masa crítica de usuarios) es utilizar información del perfil para realizar el cálculo de similitud entre usuarios. De esta manera, dos usuarios u y u' no son similares si han realizado valoraciones similares de los mismos productos, si no también si pertenecen al mismo segmento demográfico, es decir, si tienen el mismo perfil de usuario: edad, género, gustos, preferencias, educación, localización geográfica, etc. Estas técnicas construyen, por tanto, *clusters* o grupos de usuarios con una serie de características en común (de manera similar a los filtros basados en modelo).

Los métodos de recomendación que utilizan estas técnicas se conocen como filtros demográficos y clasifican a los usuarios a partir de su perfil en un conjunto de categorías. Después se utiliza la pertenencia a estas categorías para realizar las recomendaciones. Estas técnicas demográficas resultan especialmente interesantes a la hora de combinarlas con la idea de contexto para poder agrupar a los usuarios en situaciones contextuales similares y de esta forma poder sugerir contenidos multimedia en función de estas situaciones.

En la literatura se encuentran trabajos como los realizados por Krulwich [89], los cuáles desarrollan un sistema de recomendación que construye grupos demográficos partiendo de estudios de *marketing*. En otros sistemas se utilizan máquinas de aprendizaje para generar los modelos de los usuarios, este es el caso del sistema definido por Pazzani [90]. El modelo de usuario definido por estos autores utiliza el algoritmo de *Winnnow* [91] para extraer las características de los usuarios a partir de sus páginas web.

El beneficio de los métodos demográficos es que no requieren un historial muy extenso con las puntuaciones generadas por los usuarios para poder realizar recomendaciones, como ocurre con el resto de las técnicas mostradas en esta sección.

Técnicas de recomendación en redes sociales

El principal desafío de los sistemas de recomendación que trabajan sobre redes sociales radica en la necesidad de combinar información heterogénea, de fuentes diversas, de muy distinta naturaleza, tanto cualitativa como cuantitativa, y con valores comprendidos

en rangos muy dispersos. Por ese motivo, es necesario, para cada caso de aplicación, crear un modelo común de datos que permita combinar y cuantificar de manera unificada toda la información existente, así como aprovechar todos los datos que se pueden inferir de ella. Las investigaciones en este terreno son todavía incipientes, pero ya empiezan a aparecer interesantes estudios de aplicación en la literatura científica.

Orientado hacia sistemas de recomendación multimedia, Kazienko et al. [92] crean una metodología para el tratamiento de datos heterogéneos que aplican a redes sociales de compartición de imágenes o vídeos, como Flickr o Youtube. Los investigadores crean un sistema multidimensional de once capas, cada una de las cuales modela un tipo diferente de relación entre dos usuarios de una red social. Estas relaciones pueden ser, afirman, de dos tipos diferentes, dependiendo de si se establecen directamente entre usuarios o a través de algún objeto multimedia que comparten. Por lo tanto, la combinación de todas estas relaciones, situadas en las diferentes capas, da lugar a la construcción de una red multidimensional que define la relación de cada pareja de usuarios. Las once capas creadas por los investigadores para la red Flickr se basan en las listas de contactos de los usuarios (capa 1-capa 3), en las etiquetas compartidas (capa 4), en los grupos de usuarios (capa 5), en las imágenes añadidas a favoritas (capa 6-capa 8) y en las opiniones vertidas sobre imágenes creadas por los usuarios (capa 9-capa 11). Posteriormente, el sistema propuesto emplea métricas de correlación para inferir los pesos que cada una de las capas tiene para los usuarios a la hora de obtener una medida fiable de similitud entre ellos, que permite aplicar, finalmente, los algoritmos clásicos de filtrado colaborativo u otros filtros sociales, según la estructura de la Fig. 37.

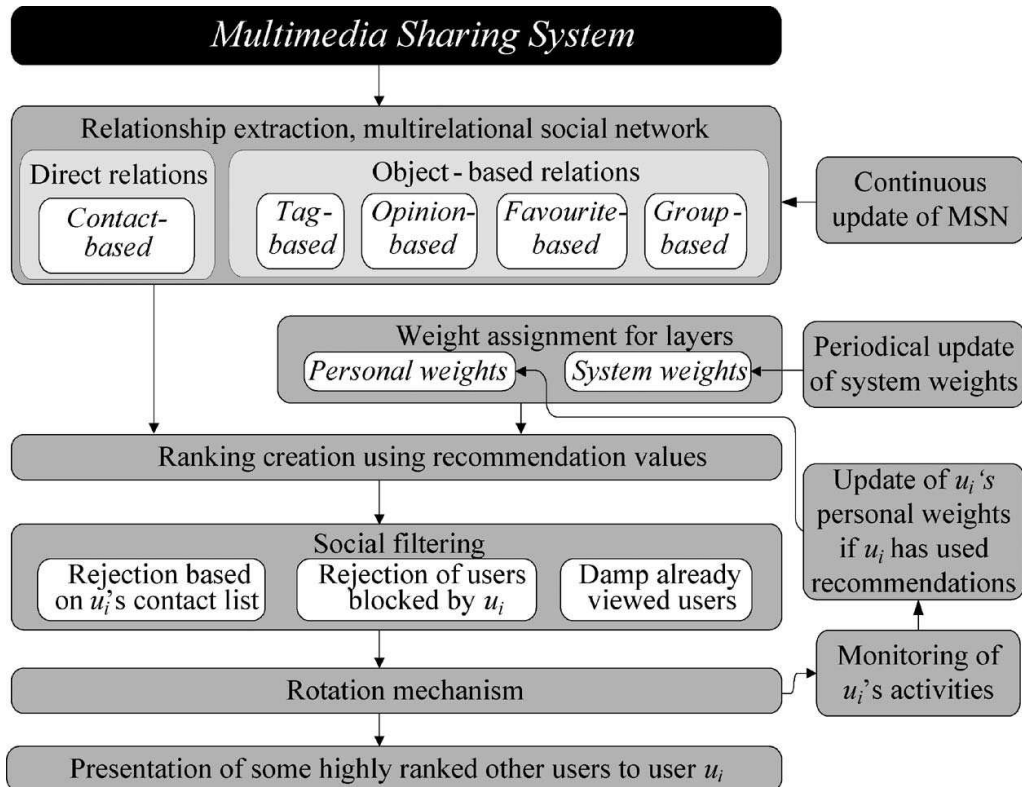


Fig. 37. Arquitectura de un sistema de recomendación para redes sociales [33]

Este modelo específico se basa, a su vez, en el modelo teórico planteado por Arazy et al. [93], en el que se consideran cuatro fuentes particulares de información, que se modelan

a su vez para obtener cuatro métricas generales (preferencias compartidas, confianza, reputación de la fuente y tie strength), que se articulan tal y como muestra la Fig. 38.

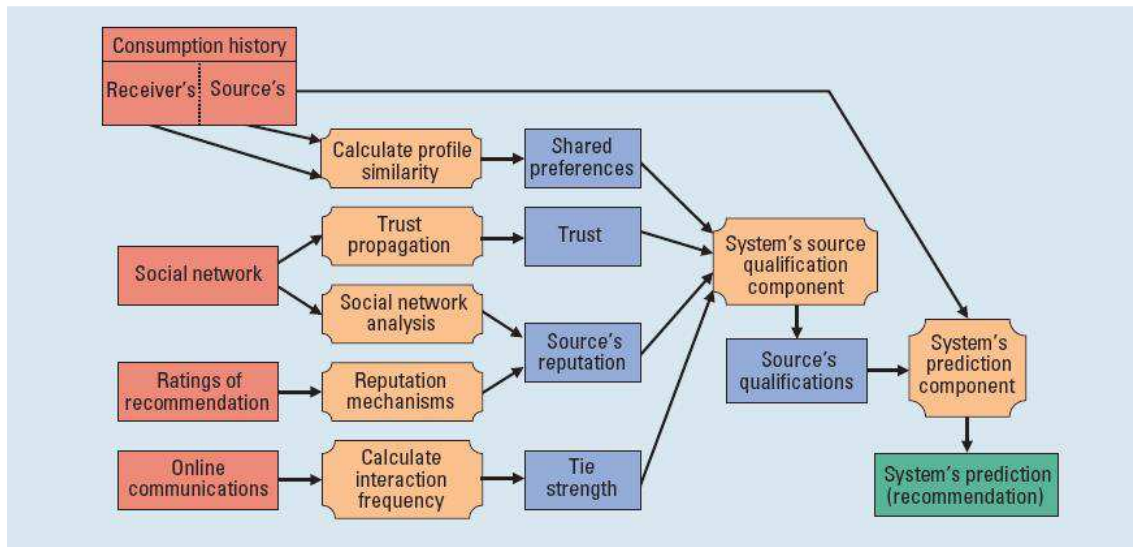


Fig. 38. Modelo de recomendación para fuentes heterogéneas en redes sociales [93]

Los investigadores también articulan este modelo sobre un sistema clásico de filtrado colaborativo. Concretamente, plantean una predicción basada en usuario a través del coeficiente de correlación de Pearson, a partir de un rating generado a partir de una suma ponderada de las cuatro fuentes de información. El problema se reduce, por lo tanto, a la óptima extracción y modelado de los datos procedentes de las redes sociales.

En este sentido, se considera fundamental el cálculo del denominado *tie strength*, atributo que fue formalizado en los años 70 y 80 por el sociólogo Granovetter [94] y en los 90 por David Krackhardt [95], y que pretende estimar la fortaleza de los lazos que unen a los miembros de una determinada comunidad. Con la irrupción de las redes sociales en Internet, las investigaciones sobre el *tie strength* han llegado a la ingeniería para aplicar las ideas sociológicas a los sistemas de información. De esta forma, se han realizado recientes estudios muy interesantes, como el que proponen Gilbert y Karahalios [96] a partir de diferentes técnicas semánticas de detección y clasificación de componentes emocionales en textos escritos. Para ello, realizan una serie de pruebas en la red social *Facebook*, donde detectan hasta 74 variables que clasifican en siete grupos: intensidad, intimidad, duración, servicios recíprocos, variables estructurales, emocionales y de distancia social. Del mismo modo, establecen una clasificación en la que ordenan las variables primarias que aportan una mayor información para el cálculo del *tie strength*.

6.2.2 Recomendación basada en contenido

Procedimientos de recomendación basada en contenido

Los sistemas de recomendación basados en contenido están estrechamente relacionados con los sistemas de recuperación y filtrado de información, ya que cualquier método de recomendación utiliza este tipo de sistemas para seleccionar y proporcionar al usuario la información recomendada, combinándolos con el elemento clave de cualquier sistema que precise de personalización: el perfil de usuario.

Como su propio nombre indica, un sistema de recomendación basado en contenido determina los ítems recomendados a partir del contenido de cada uno de ellos. Para ello, cualquier sistema de este tipo suele contar con una base de datos con dos tipos de tablas: las tablas usuario y las tablas ítem.

- En una *tabla usuario* se especifican los tipos de contenidos por los que tiene interés un usuario, siendo estos datos obtenidos en la etapa de perfilado de usuario (*perfil basado en contenido*).
- En una *tabla ítem* se encuentran los contenidos propios de un determinado ítem (*contenido del ítem*).

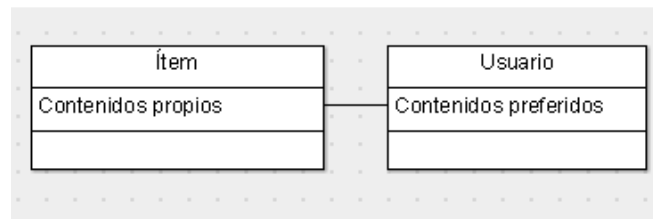


Fig. 39. Formato de tablas de la base de datos

La recomendación para cada uno de los usuarios se efectuará mediante la comparación de sus contenidos preferidos con los contenidos propios de cada uno de los ítems. Así, a partir de las relaciones de cada uno de los ítems con varios contenidos (cada ítem consta de varios contenidos propios) y de cada uno de los usuarios con dichos contenidos (cada usuario puede estar interesado en varios contenidos) se establecerán relaciones entre los ítems y los usuarios.

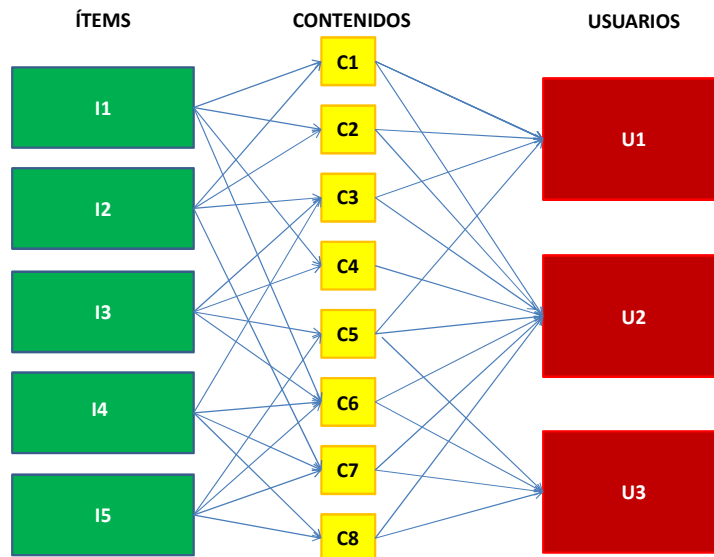


Fig. 40. Relaciones entre ítems, contenidos y usuarios

Los *contenidos propios de un ítem* son descritos en términos de *palabras clave* o *keywords*, tales como palabras pertenecientes a documentos (más frecuente en recomendación textual) o *tags* (más frecuente en recomendación audiovisual). La relevancia de estas *keywords* se mide frecuentemente mediante un parámetro denominado TF-IDF (*term frequency/inverse document frequency*) [97]. De esta manera, la relevancia de la palabra clave i en el ítem j se mediría:

$$\omega_{i,j} = TF_{i,j} \times IDF_i \quad (101)$$

Con:

$$TF_{i,j} = \frac{f_{i,j}}{\max_z f_{z,j}} \quad IDF_i = \log \frac{N}{n_i}$$

$f_{i,j}$: número de veces que la palabra clave i aparece en el ítem j

n_i : número de ítems que contienen a la palabra clave i

N : número total de ítems

Así, para el ítem j el conjunto de contenidos relevantes vendrían descritos mediante un vector de relevancias de cada una de las palabras clave.

En esta línea, también se pueden realizar aproximaciones por grafos para relacionar los contenidos entre sí según los distintos metadatos, como expresan *Peng et al. (2010)* [98] en su aproximación para recomendar fotografías según el consumo que realiza cada usuario de otras fotografías.

Asimismo, los contenidos preferidos de un usuario (perfil basado en contenido) son extraídos a partir del análisis de diferentes factores, tales como el contenido del los ítems previamente vistos y valorados por el mismo. El objetivo es obtener, al igual que en el caso anterior, un conjunto de valores (v_1, v_2, v_3, \dots) representativos de la relevancia de ciertos contenidos para el usuario. Para ello, se han utilizado distintas técnicas encaminadas a clasificar los distintos ítems en función de su relevancia, entre las que se encontrarían las siguientes:

- **Algoritmo de Rocchio [99]**: es un algoritmo de clasificación supervisado que trabaja a partir de ciertos *documentos de entrenamiento* obtenidos a partir de las valoraciones dadas a los mismos por un grupo de expertos, a los que se les proporciona un documento para establecer ellos mismos su relevancia. De esta manera, dichos expertos establecen una categoría para cada documento, por lo que una determinada categoría puede ser descrita a partir de sus documentos relevantes, a los que se les asigna un peso positivo, y de sus documentos no relevantes, con peso negativo. Así, a partir de ese entrenamiento y mediante la comparación con los documentos de entrenamiento se puede determinar la clase a la que pertenece un documento nuevo analizado.

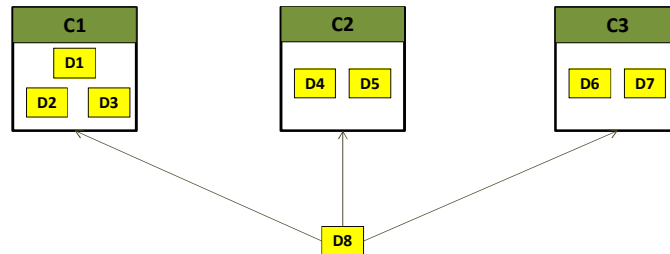


Fig. 41. Algoritmo de Rocchio

Formalmente, el algoritmo se puede enunciar de la siguiente manera:

$$\vec{Q}_m = (a * \vec{Q}_0) + \left(b * \frac{1}{|D_r|} * \sum_{\vec{D}_j \in D_r} \vec{D}_j \right) - \left(c * \frac{1}{|D_{nr}|} * \sum_{\vec{D}_k \in D_{nr}} \vec{D}_k \right) \quad (102)$$

- 1) \vec{Q}_m es el vector final que establece el peso de cada uno de los rasgos en relación a una clase.
 - 2) \vec{Q}_0 es el vector inicial que establece el peso de cada uno de los rasgos en relación a una clase.
 - 3) D_r es el conjunto de documentos relevantes.
 - 4) D_{nr} es el conjunto de documentos no relevantes.
 - 5) a es el factor de inicialización.
 - 6) b es el factor de relevancia.
 - 7) c es el factor de no relevancia.
- **Algoritmo de Winnow [91]:** es una técnica de aprendizaje en máquinas que implementa un clasificador lineal a partir de ejemplos etiquetados. Se utiliza para identificar rasgos relevantes de un documento entre un amplio rango de ellos. En una clasificación binaria (relevante/no relevante) se utilizan tres parámetros [103]:
 - 1) Umbral θ : normalmente vale 1, y sirve de referencia al inicializar los pesos de todos los rasgos.
 - 2) Parámetro de promoción $\alpha > 1$
 - 3) Parámetro de degradación $1 > \beta > 0$

Los documentos y las clases se representan mediante vectores $d = \{d_1, d_2, d_3, d_4, d_5, \dots\}$ y $c = \{c_1, c_2, c_3, c_4, c_5, \dots\}$ donde cada d_i representa la importancia (peso) del rasgo i en el documento y cada c_i representa, análogamente, la importancia del rasgo i en la clase. El algoritmo de entrenamiento se realiza de la siguiente forma:

- 1) Si $\sum_i d_i c_i > \theta$ predice 1, si no predice 0, siendo d_i el peso del rasgo i en el documento ejemplo y c_i el peso del rasgo i en la clase.
 - 2) Si predice 0 cuando el documento está etiquetado como 1, los valores c_i que predijeron 1 se multiplican por α .
 - 3) Si predice 1 cuando el documento está etiquetado como 0, los valores c_i que predijeron 0 se multiplican por β .
- **Clasificador Bayesiano [100]:** en este caso el proceso de clasificación está basado en la teoría de la probabilidad, ya que se aplica el teorema de Bayes para calcular la probabilidad de pertenencia de un rasgo a una categoría, obteniéndose la clasificación más probable. El algoritmo de este tipo más simple y conocido es el denominado *naive Bayes*. Dado un conjunto X de variables independientes $X = \{x_1, x_2, x_3, x_4, x_5, \dots\}$ y un conjunto de categorías $C = \{c_1, c_2, c_3, c_4, c_5, \dots\}$, siguiendo la regla de Bayes, la probabilidad de que X pertenezca a la categoría c_i cumple la siguiente proporcionalidad:

$$p(c_i/x_1, x_2, x_3, \dots, x_N) \propto p(x_1, x_2, x_3, \dots, x_N/c_i)p(c_i) \quad (103)$$

Teniendo en cuenta que las variables x_i son independientes lo anterior puede ser reescrito, concretamente:

$$p(c_i/X) = \frac{p(c_i) \prod_{k=1}^N p(x_k/c_i)}{p(X)} \quad (104)$$

Una vez calculadas las probabilidades para todas las categorías, se selecciona aquella con una probabilidad mayor.

Con este algoritmo se pueden clasificar ítems a partir del conocimiento de las categorías a las que pertenecen sus rasgos.

Una vez que se han sido obtenidos el perfil basado en contenido y el contenido del ítem se procede al cálculo de la función *utilidad*, a partir de la cual se determinará la recomendación. Para su cálculo se pueden emplear distintos procedimientos, tales como la *similitud por coseno* [101] de los dos parámetros, así como distintos métodos heurísticos, tales como *clustering*, redes neuronales o árboles de decisión [102]:

- **Clustering:** es una técnica de clasificación no supervisada (no utiliza vectores de entrenamiento) que consiste en la agrupación de elementos según un determinado parámetro, teniendo en cuenta que los elementos de parámetro similar se encontrarán dentro del mismo *cluster*. En [104] se desarrolla un sistema de recomendación de aplicaciones de dispositivos móviles según el contexto y el patrón de uso basado en *clustering*. De la misma manera, en [105] se propone un sistema de recomendación de piezas de música mediante *clustering*.
- **Redes neuronales:** las redes neuronales basadas en mapas topológicos autorganizativos (*Self Organizing Maps, SOM*) [106] implementan técnicas de clasificación. También implementan algoritmos no supervisados, por lo que son muy apropiadas cuando no se tienen a priori datos para realizar un entrenamiento. Han sido muy utilizadas en clasificación de documentos, tales como en el proyecto *WEBSOM* [107].
- **Árboles de decisión:** esta técnica se basa en predicciones de tipo lógicas, normalmente utilizando datos almacenados. Es muy similar a la recomendación basada en reglas. En [108] y [109] se proponen sistemas de recomendación de películas basados en árboles de decisión, a partir de los géneros de las mismas.

Técnicas de recomendación visual

Como cualquier técnica de recomendación basada en contenido, en la recomendación de contenido visual se precisa del conocimiento del contenido de interés para el usuario, que será esencial para realizar las comparaciones pertinentes con el material que se quiere recomendar. Este contenido será representado mediante distintos descriptores, que deberán estar correlacionados con la información que se almacena de los usuarios, y que se obtiene implícita o explícitamente.

Debido a las diferentes características, formatos y utilidades del material visual, es necesario realizar una clasificación del mismo según su dinamismo (material visual estático y dinámico) y tratar los dos tipos por separado. Considerando esto, se establecerá una clasificación entre imágenes (estático) y vídeo (dinámico).

a) Imágenes

Las técnicas de recuperación de imágenes cobran una gran importancia en las tareas de recomendación visual, ya que un elemento crucial en cualquier sistema de

recomendación es la selección y recuperación del contenido que se va a recomendar de acuerdo a determinados parámetros. Estos parámetros deben ser cuidadosamente elegidos, ya que determinarán dicha recuperación. Las dos principales técnicas de recuperación de imágenes son las siguientes:

1. *Meta-búsqueda de imágenes*: búsqueda basada en metadatos asociados a las imágenes, como palabras clave o texto.
2. *Basada en contenido (CBIR)*: recuperación de imágenes a través de similitud visual con otra imagen proporcionada por el usuario, o por contener determinadas propiedades especificadas por el mismo.

CBIR es la técnica más comúnmente utilizada. Provee imágenes a partir del análisis de rasgos de una determinada imagen anteriormente proporcionada. Los rasgos más utilizados son el color, la textura y la forma. En un sistema de recomendación, CBIR debe aprender de las preferencias del usuario para que en una posterior iteración se produzca una mejora en las características presentadas. Este proceso de aprendizaje se denomina *preference feedback* o *relevance feedback* [113]. En [114] se desarrolla un sistema que implementa *relevance feedback* tratando de establecer una relación entre factores de bajo y de alto nivel, así como resolver el problema de la subjetividad humana frente a la manera de actuar de las máquinas. Esto se realiza mediante una red en la que los distintos ítems están relacionados y sus relaciones están ponderadas, de manera que dichas ponderaciones se vayan actualizando en cada fase de realimentación. Asimismo, la empresa Daedalus ha desarrollado un sistema de CBIR para la recuperación de imágenes pictóricas [110].

Entre las características que pueden ser extraídas de las imágenes, se encuentran las definidas en [112]. Sin embargo, este tema ya ha sido ampliamente tratado en el capítulo 5 de esta tesis.

Una vez que se cuenta con las características definidas, el siguiente paso es el cálculo de las diferencias entre distintas imágenes, para lo que se utilizan distintos tipos de distancias, entre las que se encuentran las siguientes, definidas en el Diccionario de distancias [111]:

- Euclídea

$$d(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (105)$$

- Mahalanobis

$$d_m(\vec{x}, \vec{y}) = \sqrt{(\vec{x} - \vec{y})^T \Sigma (\vec{x} - \vec{y})}^{-1} \quad (106)$$

- Hausdorff

$$d_H(X, Y) = \max \{ \sup_{x \in X} \inf_{y \in Y} |xy|, \sup_{y \in Y} \inf_{x \in X} |xy| \} \quad (107)$$

- Minkowski

$$d_m(\vec{x}, \vec{y}) = (\sum_{i=1}^n |x_i - y_i|^p)^{1/p} \quad (108)$$

- Manhattan

$$d(x, y) = |x_2 - x_1| + |y_2 - y_1| \quad (109)$$

- Chebyshev

$$d(x, y) = \max(|x_2 - x_1|, |y_2 - y_1|) \quad (110)$$

La distancia de Minkowski es una generalización de las distancias Euclídea, Manhattan y Chebyshev (o Chebychev), donde un parámetro p debe ser definido. Si $p = 1$, es la distancia de Manhattan, si $p = 2$, es la distancia Euclídea y, finalmente si $p = \infty$, es la distancia de Chebyshev. Adicionalmente, la distancia Euclídea es un caso particular de la distancia de Mahalanobis: en la distancia Euclídea no se tiene en cuenta la correlación entre los atributos (dimensiones). Estas son las distancias más utilizadas para los distintos tipos de sistemas de recomendación. Sin embargo, en algunos casos, en los que se consideran distancias a conjunto de elementos, se emplea la distancia Hausdorff.

Estas distancias son aplicables tanto a imágenes como a vídeos, ya que se aplican a la obtención del grado de disimilaridad entre dos elementos parametrizados, independientemente de su naturaleza.

En un sistema típico de recomendación basada en contenido los contenidos a recomendar corresponderán a aquellos cuyas distancias al ítem objetivo (ítem del que se conoce que es de interés para el usuario) sean menores. Estas distancias pueden ser calculadas directamente a partir del ítem objetivo y de todas las imágenes con las que se cuente para recomendar, o bien se pueden utilizar determinadas técnicas para facilitar el proceso. En [115] se propone una técnica para la recomendación basada en contenido que utiliza dos parámetros basados en las características de las imágenes:

- *Ítems equivalentes*: si dos imágenes tienen rasgos muy similares son consideradas equivalentes. Por ejemplo, dos fotografías de un mismo edificio pero con claridad distinta. Estas imágenes producirán un efecto parecido en usuarios con gustos similares.
- *Usuarios virtuales*: un usuario virtual es un conjunto de imágenes con un determinado rasgo en común que puede ser relevante en las preferencias de un usuario. Dicho rasgo puede referirse a la forma, el color, la textura o el diseñador. Mediante el uso de usuarios virtuales, se puede modelar una recomendación para usuarios con gustos similares a los recogidos en los usuarios virtuales.

Otro de los parámetros que se pueden utilizar para facilitar la recomendación es el contexto. En [116] se propone un sistema de recomendación de imágenes basado en contenido y en contexto, que modela los perfiles de los usuarios en función de las

imágenes preferidas y del contexto al que pertenecen, como por ejemplo, el día de la semana en el que suelen buscar determinadas imágenes.

b) Vídeos

El contenido perteneciente a un vídeo presenta muchas más complejidades a la hora de tratar y analizar que una imagen. En los sistemas de recomendación de vídeo basado en contenido es preciso contar con la mayor cantidad de información posible sobre los mismos. Para realizar la recomendación se deben conocer parámetros con distintas propiedades, que podrán ser clasificados según tres niveles de clasificación: nivel cognitivo, nivel afectivo y nivel contextual, tal y como indica la siguiente figura.

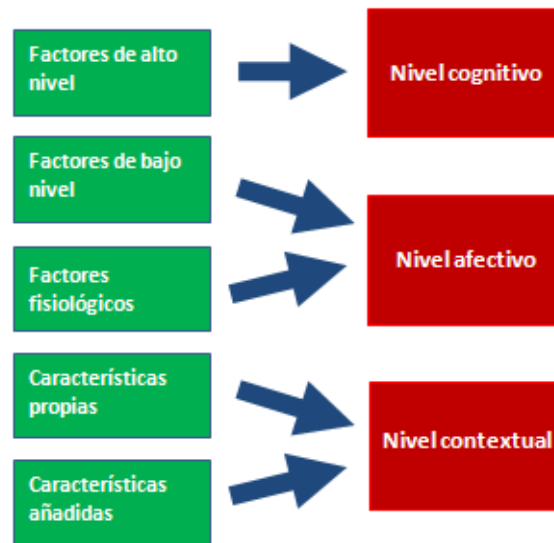


Fig. 42. Niveles de clasificación de vídeos

- **Nivel cognitivo**

El nivel cognitivo representa la información, que contiene un vídeo, relativa a factores de alto nivel, tales como el texto y las imágenes contenidas en el mismo, que pueden ser reconocidas y diferenciadas. Por ejemplo, podría proveer información sobre si en un vídeo aparecen dos personas hablando, una panorámica de Nueva York, un cartel, una carrera de coches, etc. Es decir, representa, los *hechos* contenidos en el vídeo. Los datos necesarios para la implementación de este nivel se obtendrán en la tarea 4.2, en la que se desarrollarán distintos mecanismos de detección de contenidos en vídeos, por ejemplo mecanismos de detección de caras. Estos datos serán proporcionados en forma de *palabras clave* representativas del contenido de la imagen o del vídeo.

Los datos semánticos extraídos son la base para la configuración de la recomendación. El primer paso a realizar es la clasificación y agrupación de los distintos contenidos según su temática, lo que se realiza calculando similitudes semánticas entre los mismos. Para ello se suele utilizar el parámetro denominado *tf-idf* (*term frequency–inverse document frequency*). Este parámetro es comúnmente utilizado en sistemas de recuperación de textos, ya que mide la frecuencia con la que determinadas palabras aparecen en un documento. En el tratamiento de contenido visual, también es factible la utilización de este parámetro, debido a que permite medir la frecuencia con la que aparecen los tags en un ítem de naturaleza visual. Una vez que se han obtenido dichas

frecuencias, es posible calcular las diferencias semánticas para su posterior uso en la recomendación, lo que ha sido realizado mediante distintas técnicas:

- **Bayes:** en [121] se obtienen similitudes semánticas entre las palabras clave en función del número de ítems en los que coincidan. Estas coincidencias de palabras se utilizan como datos de entrenamiento en una red Bayesiana, de tal forma que sea posible relacionar directamente las palabras clave pertenecientes a ítems de interés para el usuario con otras de otros ítems. Asimismo, en [123] se desarrolla un recomendador de noticias a partir de la base de datos de la CNN y de las palabras claves contenidas en ella. Los vídeos son evaluados por los usuarios de forma binaria (valoración positiva y negativa) y se relaciona los votos recibidos por los vídeos con votos de las palabras clave mediante un clasificador bayesiano.
- **Coseno y ontología:** en [122] se calculan similitudes entre las palabras clave mediante la similitud por coseno y mediante una ontología con la que se cuenta para la realización de similitudes semánticas.
- **Spread activation:** en [124] se realizan asociaciones semánticas entre las palabras clave y se calculan las similitudes mediante el método de *spread activation*, que consiste en la búsqueda a través de una red mediante ponderación de los enlaces entre cada uno de sus nodos. La activación asociada a cada palabra es la suma de las activaciones de sus enlaces de entrada, ponderada por el factor de importancia de cada enlace.
- **Vecinos cercanos:** consiste en el cálculo de las distancias del ítem objetivo a todos los ítems disponibles y en la selección de aquéllos con menor distancia. En [126] se realiza una descomposición del vídeo en segmentos basándose en la homogeneidad del contenido. Teniendo en cuenta que un vídeo está dividido en planos y que cada plano a su vez está compuesto por *frames*, se seleccionan los *frames* vecinos que contienen una escena similar. Cada una de estas escenas comunes, compuestas por uno o varios *frames*, son las consideradas como *segmentos de vídeo*. Las distancias que se calculan son las aquellas comprendidas entre el ítem objetivo y cada uno de dichos segmentos de vídeo, obteniéndose los vecinos más cercanos de dos formas: o bien seleccionando los K cuya distancia sea menor, o bien estableciendo un umbral como distancia máxima y eliminando todos los elementos que estén a una distancia mayor.

- **Nivel afectivo**

En el nivel afectivo están contenidas las posibles reacciones que puede producir un vídeo en un usuario. Esto puede ser obtenido mediante el análisis de ciertos factores a bajo nivel o incluso directamente mediante estudio de factores fisiológicos del usuario. En [117] se obtienen distintos factores de influencia en un usuario mediante características como la intensidad del sonido, la velocidad de los cambios de plano, o el ritmo. Asimismo, en [118] se realiza un análisis de las reacciones del usuario mediante mediciones de ciertas características fisiológicas, tales como la presión sanguínea, los latidos del corazón o la resistencia de la piel.

- **Nivel contextual**

El nivel contextual se construye a partir de las características del vídeo útiles para realizar comparaciones con otros vídeos. Estas características se dividen en dos tipos:

- **Características propias:** son los descriptores que sitúan al vídeo en un entorno espacio-temporal, y que generalmente son proporcionadas por los metadatos. Por ejemplo, título, descripción, género, año, lugar... En [119] se realiza una extracción de este tipo de características a través de técnicas de resumen de vídeo, y se realiza una recomendación mediante la comparación de las características extraídas con el perfil del usuario.
- **Características añadidas:** en el caso de que el vídeo haya sido publicado en un determinado sitio web, en ocasiones es posible acceder a este tipo de características relativas a la experiencia de los usuarios, tales como opiniones, votaciones o número de visionados. En [125] se realiza una caracterización de los vídeos a partir de los tags asignados por los distintos usuarios, y se construyen perfiles de usuarios en función de los tags que ha utilizado cada uno de ellos. De esta manera, se calculan similitudes entre los mismos mediante similitud de coseno. Es una técnica muy parecida al filtrado colaborativo, sólo que en este caso no se utilizan las votaciones de los usuarios, sino los tags, que están relacionados con el contenido. Así, un usuario es similar a otro si coinciden en los vídeos visionados y en los tags asignados a los mismos.

En [119] se desarrolla un sistema de recomendación basado en información contextual y de contenido. Para ello se construye un grafo contextual en el que los nodos son los vídeos (cada uno de ellos con características propias), y cada una de las aristas está ponderada según la relación entre los dos vídeos que conecta, calculándose dicha ponderación a partir de las características propias y añadidas.

6.2.3 Algoritmos de hibridación

Análisis de algoritmos monomodales

Los filtros colaborativos son en general las técnicas de recomendación más habituales en la construcción de sistemas de recomendación para servicios electrónicos. Sin embargo, dependen del solapamiento o coincidencia en las valoraciones de los usuarios y encuentran dificultades de precisión cuando el conjunto de valoraciones es disperso y escaso, es decir, cuando sólo unos pocos usuarios han valorado los mismos productos. Aunque los filtros basados en modelos han mejorado su rendimiento ante la dispersión de datos, el filtrado colaborativo sigue teniendo grandes problemas cuando se encuentran con nuevos usuarios y nuevos recursos. En cualquier caso, hay que tener en cuenta que las técnicas colaborativas son las más adecuadas cuando existe un conjunto de usuarios lo suficientemente grande y participativo que proporcionan evidencias de sus intereses a partir de las puntuaciones (o sencillamente a través de sus adquisiciones de productos). Cuando este universo de ofertas cambia demasiado rápido (por ejemplo, en el caso de noticias de un periódico) o la participación es muy baja, la precisión de las técnicas colaborativas se ve seriamente afectada. Una aproximación interesante para limitar el efecto de la baja información sobre los usuarios es realizar un análisis demográfico, que permite eliminar el problema de los nuevos usuarios. El problema habitual de estos sistemas es la reticencia cada vez mayor de los usuarios a proporcionar información de carácter personal, teniendo en cuenta que los datos más pertinentes para la recomendación son posiblemente aquellos que el usuario está menos dispuesto a compartir.

Las técnicas basadas en contenido también tienen el mismo problema para nuevos productos y usuarios. Necesitan acumular la suficiente evidencia, es decir, *ratings*, para poder construir un clasificador que genere recomendaciones adecuadas. Además, los algoritmos basados en contenido sólo pueden trabajar con los datos disponibles sobre la descripción del recurso, por ejemplo, en el caso de una película: el título, los actores, en qué año se rodó, temática, etc. Aunque esta información es fundamental para generar recomendaciones precisas que satisfagan las demandas de los usuarios, limita la capacidad de estos algoritmos para recomendar más allá de la descripción de los recursos. En este sentido, usuarios a los que les guste la música *jazz* también pueden estar interesados en música de fusión. Los filtros colaborativos serían capaces de generar recomendaciones en esta dirección, ya que no dependen de las características de los productos, sino de las valoraciones y opiniones de los usuarios. En cambio, las técnicas basadas en contenido entrenadas en las preferencias de un usuario sobre el estilo *jazz* serían ciegas a la música de fusión, sobre todo, si ambos estilos musicales no comparten atributos de descripción. Sin embargo, existe una interesante línea alrededor de la extracción y modelado de características de bajo nivel, ya que se pueden encontrar relaciones estéticas o estructurales que no son visibles a simple vista, y que aportan una mayor variedad y riqueza en las recomendaciones. Esta línea, por lo tanto, puede suponer un gran complemento a las recomendaciones basada en tags y al filtrado colaborativo.

Esta misma limitación la encontramos en los algoritmos basados en conocimiento. Sin embargo, estas técnicas no se ven afectadas, como las dos anteriores, por la dispersión o escasez de evidencia sobre las valoraciones de los productos por parte de los recursos. La potencia de estas técnicas se encuentra en que permiten utilizar cualquier atributo que se considere necesario de los productos a la hora de calcular su utilidad para un usuario. Por tanto, la introducción de un nuevo producto o la llegada de un nuevo usuario al sistema no merma la eficiencia de estos algoritmos. Sin embargo, se ven claramente limitados al trabajar exclusivamente con la descripción de los recursos y el perfil de los usuarios. No pueden, como es el caso de los filtros colaborativos, sugerir recursos teniendo en cuenta el comportamiento global del resto de usuarios del sistema. Estos algoritmos pueden ser muy precisos cuando el usuario tiene muy claros sus intereses y los expresa de manera correcta de acuerdo con la formalización del algoritmo y la representación de los recursos. Sin embargo, si el usuario no conoce demasiado el dominio, tiene demasiadas dudas acerca de sus intereses o sencillamente no quiere tener que formular explícitamente su demanda, entonces los algoritmos basados en conocimiento pierden toda su capacidad para generar recomendaciones relevantes para los usuarios.

Metodologías de hibridación

El filtrado híbrido aprovecha las ventajas de otros algoritmos de recomendación, especialmente de los basados en contenido y los colaborativos, combinando sus predicciones, por lo que utiliza tanto las apreciaciones de los ítems como las descripciones de las preferencias de los usuarios.

Hay diferentes formas de combinar ambos sistemas [7]:

- 1) *Implementar por separado los sistemas basados en contenido y los colaborativos y combinar las predicciones.*

En este caso existen dos posibles escenarios. En el primero, se pueden combinar las valoraciones obtenidas de sistemas de recomendación individual en una recomendación final utilizando una combinación lineal de valoraciones [127] o un esquema de votación [128]. Alternativamente, se puede emplear en cualquier momento uno de los recomendadores individuales, eligiendo el que sea mejor según una recomendación métrica de calidad [129].

- 2) *Incorporar ciertas características de los sistemas basados en contenido en un sistema colaborativo.*

Algunos sistemas de recomendación híbridos, incluidos Fab [73] y la aproximación de “colaboración a través de contenido” descrita en [127], se basan en técnicas colaborativas tradicionales, pero mantienen los perfiles basados en contenido de cada usuario. Estos perfiles se utilizan habitualmente para calcular la similitud entre dos usuarios. Esto permite solucionar algunos problemas de dispersión de un sistema puramente colaborativo, dado que una pareja de usuarios no tiene un gran número de valoraciones en común. Otra ventaja de esta aproximación radica en que para un usuario no sólo habría recomendaciones de ítems que hubieran sido altamente valorados por usuarios con perfiles similares, sino también directamente cuando la descripción del elemento concuerda claramente con el perfil del usuario [73]. Otros sistemas emplean una aproximación similar usando los llamados *filterbots*: [130] agentes de análisis de contenido especializados que actúan como participantes adicionales en el filtrado colaborativo. Como resultado, los usuarios cuyas valoraciones son similares a las de algunos *filterbots* podrán recibir recomendaciones de ellos. De una forma similar, otros sistemas [131] utilizan una aproximación colaborativa en el que el tradicional vector de valoraciones del usuario está aumentado con valoraciones adicionales, calculadas usando un predictor basado en contenido.

- 3) *Incorporar ciertas características de los sistemas colaborativos en un sistema basado en contenido.*

La aproximación más popular de esta categoría consiste en utilizar una técnica de reducción de la dimensionalidad en un grupo de perfiles generados por un sistema basado en contenido. Por ejemplo, en [136] se utiliza un indexado semántico latente (LSI) para crear una visión colaborativa de una colección de perfiles de usuario, en el que estos perfiles se representan mediante “*term vectors*”, mejorando el rendimiento respecto al basado puramente en contenido.

- 4) *Construir un modelo único general que incorpore características de los sistemas basados en contenido y colaborativos.*

Muchos investigadores han seguido esta aproximación en los últimos años. Por ejemplo, algunas líneas [139] proponen utilizar características basadas en contenido y colaborativas (como la edad o género de los usuarios o el género

de las películas) en un clasificador único basado en reglas. Otras opciones propuestas [140] consisten en un método probabilístico unificado para combinar recomendaciones colaborativas y basadas en contenido, basándose en un análisis semántico probabilístico [141]. Finalmente, otras aproximaciones [142][143] utilizan modelos de regresión bayesiana que emplean cadenas de Markov y métodos de Monte Carlo para estimar y predecir parámetros.

Algoritmos de hibridación

Las técnicas de recomendación analizadas no funcionan en todas las situaciones, ya que cada una presenta sus restricciones particulares. Por ejemplo, en el caso de los filtros colaborativos, no disponer de una masa crítica de usuarios y un volumen elevado de valoraciones de los recursos implica que las sugerencias que se obtienen no sean del todo precisas y fiables. Para superar estas limitaciones se pueden combinar estos métodos, dando lugar a técnicas híbridas de recomendación. Para ello, en la literatura encontramos varias maneras formas de combinar tanto las descripciones de los elementos que se pueden recomendar como los resultados obtenidos a través de diferentes métodos. Algunos de los mecanismos para combinar técnicas de recomendación son los siguientes:

- **Weighted.** Esta técnica de recomendación se basa en la combinación de la puntuación obtenida por medio de los distintos metodos de recomendación. Esta combinación puede ser aplicada utilizando una función que pondera la importancia de cada uno de los recomendadores. En este grupo encontramos el sistema P-Tango de Claypool [127], que usa un sistema híbrido de filtrado colaborativo y basado en el contenido que va calibrando el peso de cada parte en función de las puntuaciones que los usuarios dan a las recomendaciones. Encontramos también el trabajo de Pazzani [90] que combina el filtrado colaborativo, el basado en contenido y el demográfico. Este sistema no utiliza una puntuación numérica sino que los resultados de cada método se interpretan como un conjunto de votos que se combinan utilizando un esquema consensuado.
- **Switching.** Esta técnica selecciona el recomendador más adecuado para cada tarea analizando los resultados obtenidos en evidencia pasada. Para realizar esta tarea es necesario especificar las estrategias de selección que se van a utilizar. El sistema *DailyLearner* usa una combinación de recomendación basada en el contenido y de filtrado colaborativo. Primero utiliza el método basado en el contenido y si en resultado obtenido no es una recomendación lo suficientemente buena lo intenta utilizando el método basado en filtrado colaborativo. La utilización del filtrado colaborativo posibilita al recomendador realizar recomendaciones que no estén cerradas por la semántica, es decir, que se pueden ofrecer recursos que no tienen propiedades en común con la demanda. Cohen [136] propone un método híbrido más directo ya que utiliza la confirmación del usuario de las recomendaciones obtenidas.
- **Mixed.** Esta técnica permite la utilización de diferentes técnicas de recomendación mezclando los resultados obtenidos por cada una de los recomendadores. Se le presentan al usuario como una única lista de sugerencias. El sistema PTV de Smith y Cotter [133] es usado para recomendar programas de

televisión. Usa un método híbrido basado en contenido y filtrado colaborativo construido sobre las preferencias del usuario. Las recomendaciones obtenidas por los dos métodos son mezcladas para conseguir una única lista de recomendaciones. Otros ejemplos de este tipo de recomendadores son los del ProfBuilder de Wasfi [134] y PickAFlick de Burkner *et al.* [135].

- **Feature combination.** Esta técnica utiliza características obtenidas de diferentes fuentes de información en un único algoritmo, por lo que la combinación se realiza al nivel de los datos que se utilizan para la recomendación. Dentro de este grupo encontramos el trabajo de Basu *et al.* [136] los cuales comentan un experimento en el que se induce el aprendizaje de reglas, aplicado para la recomendación de películas usando las puntuaciones de los usuarios y las características del contenido. Mostraban mejoras significativas sobre un sistema de recomendación únicamente colaborativo.
- **Cascade.** Esta técnica utiliza los resultados obtenidos por un recomendador para la entrada del siguiente. De este modo, los resultados son redefinidos en la ejecución de cada recomendador. El recomendador de restaurantes *Entree* usa una combinación de métodos basados en el conocimiento y filtrado colaborativo. Utiliza el primero de ellos para hacer las recomendaciones y utiliza el segundo método en el caso de que se encuentre un empate entre las puntuaciones obtenidos por el primer método.
- **Feature augmentation.** Técnica que utiliza los resultados obtenidos por una técnica de recomendación como entrada para la siguiente. Por ejemplo, encontramos el sistema *Libra* [137] que hace recomendaciones basadas en el contenido de los libros de Amazon.com, usando un clasificador de texto basado en redes bayesianas. En los textos que conforman el *corpus* se incluyen los autores y títulos relacionados que aporta el filtrado colaborativo y que incluye el propio portal de Amazon.com. El grupo de investigación GroupLens trabaja con el filtrado de noticias Usenet empleando también *feature augmentation*. Este grupo implementa una serie de "*filterbots*" que utilizan técnicas basadas en el conocimiento usando criterios específicos. Estos *bots* contribuyen a la puntuación de las noticias al igual que las valoraciones de los usuarios.
- **Meta-level.** Técnica que utiliza un modelo aprendido a través de una técnica como entrada para el siguiente método de recomendación. En esta línea, encontramos el trabajo realizado por Pazzani [90], que propone una recomendación colaborativa a través del contenido. También encontramos el recomendador de restaurante propuesto por Littlestone y Warmuth [138], el cual describe para cada usuario las características que se deben usar para realizar sus recomendaciones. Condliff *et al.* [142] usa dos niveles de redes bayesianas. La primera se utiliza para la clasificación basada en el contenido de cada uno de los usuarios. Después se crean parametros de aprendizaje para construir perfiles de usuario que puedan ser utilizados por un método colaborativo.

6.3 Filtrado colaborativo

Como se ha comentado en la introducción del capítulo, la implementación de un algoritmo de filtrado colaborativo dentro de esta tesis corresponde a las necesidades de hibridación de un sistema de recomendación basado únicamente en contenido y, más especialmente, el basado en características formales y estéticas de los textos audiovisuales. Por esa razón, se diseña e implementa un algoritmo de filtrado colaborativo adaptado al caso de uso del corpus siguiendo los modelos expuestos en el estado del arte. Existen posibilidades más potentes del filtrado colaborativo que tienen en cuenta más cantidad de información, como los instantes temporales de adquisición de los *ratings* de los usuarios, para lo que se construyen modelos probabilístico adecuados. Sin embargo, la investigación en técnicas de filtrado colaborativo queda fuera del alcance de esta tesis, y se emplea únicamente a nivel conceptual para la realización de pruebas de hibridación y para fijar un marco de referencia asentado en la comunidad científica con el que comparar los resultados obtenidos.

El algoritmo parte de los mismos datos que utiliza el recomendador basado en contenido, pero utiliza únicamente las valoraciones de los usuarios sobre los ítems. No precisa la información sobre el contenido de los ítems. Estos datos están recogidos en el marco de pruebas ya descrito en el apartado 5 del capítulo 5 de esta tesis. Por lo tanto, se utilizará únicamente el identificador del usuario, el del ítem, y el *rating* (preferencia o grado de interés) de cada usuario para cada ítem.

Se pueden diferenciar dos etapas dentro del algoritmo: la de **predicción**, que es el filtrado colaborativo propiamente dicho, y la de **recomendación**.

La primera etapa recibe como dato de entrada en cada ejecución los identificadores de un usuario y un ítem, y debe sacar como resultado la predicción del rating (gusto, interés) que ese usuario le daría al ítem en el caso de que lo consumiera. El primer paso, como es lógico, consiste en comprobar si el elemento ya está consumido; en ese caso, se devuelve directamente ese valor de interés almacenado en la base de datos. Sin embargo, el caso normal es que la predicción se solicite para un elemento que no ha sido consumido.

El proceso de predicción, a su vez, se divide en dos etapas, tal y como ya ha sido comentado en el estado del arte, con lo que se sigue el procedimiento básico. Dentro de las opciones posibles, se decide utilizar un acercamiento basado en usuario.

La primera etapa es independiente del ítem para el que se quiere realizar la predicción, ya que únicamente consiste en crear un *cluster* de los usuarios afines (o vecindario) al usuario para el que se realiza el filtrado, otorgando un peso a cada uno de ellos, que corresponderá con la similitud o grado de afinidad. Por esta razón, esta subetapa puede tener lugar en un proceso offline, ya que se basa en datos previamente almacenados, sin necesitar ninguna petición ni información suplementaria del usuario. El proceso de formación del *cluster* se realiza calculando la similitud del usuario con cada uno de los demás usuarios existentes en la base de datos, lo que se obtiene utilizando la fórmula del coeficiente de correlación de Pearson, ya comentado en el estado del arte pero que se repite a continuación por su importancia.

El resultado de aplicar la fórmula de Pearson es el grado de similitud, que servirá para otorgar un peso a cada uno de los usuarios, siendo los N primeros los que formarán el *cluster* a partir del cual se calcula posteriormente la predicción.

El tamaño del *cluster*, N , es configurable, y los resultados finales dependen de su dimensionamiento. La decisión sobre el valor de N , por un lado, debe tomarse teniendo en cuenta las capacidades computacionales del servidor y, por otro, está directamente relacionada con la cantidad y dispersión de datos almacenados, tanto el número de usuarios como el número de ítems posibles y la cantidad de *ratings* de cada usuario. Por ejemplo, una base de datos con muchos ítems y pocos ratings de cada usuario obligaría a crear *clusters* de un tamaño bastante grande debido a que, de lo contrario, a la hora de realizar una predicción para un ítem, sería muy difícil encontrar dentro del *cluster* otros usuarios que hayan consumido ese ítem y para los que, por lo tanto, existan datos que utilizar para que el algoritmo pueda ofrecer resultados aceptables. Por otro lado, si muchos usuarios han consumido (por lo tanto existe información) la mayoría de los ítems, es necesario tener cuidado y fijar un tamaño de *cluster* limitado, ya que de lo contrario, y a pesar de la ponderación por pesos, la entrada de usuarios poco afines podría distorsionar la predicción, cuando bastaría con tener en cuenta la información de los que realmente son más próximos. Dado el corpus de prueba, se fijó como aceptable un tamaño de $N=20$.

En la Fig. 43 se presenta un ejemplo, en el que se ha elegido un *cluster* de 5 usuarios para obtener los vecinos del usuario A. Cada vértice representa un usuario, y las aristas la similitud entre usuarios (en este caso, entre el usuario A y cada uno de los demás). Mayor distancia implica menor similitud.

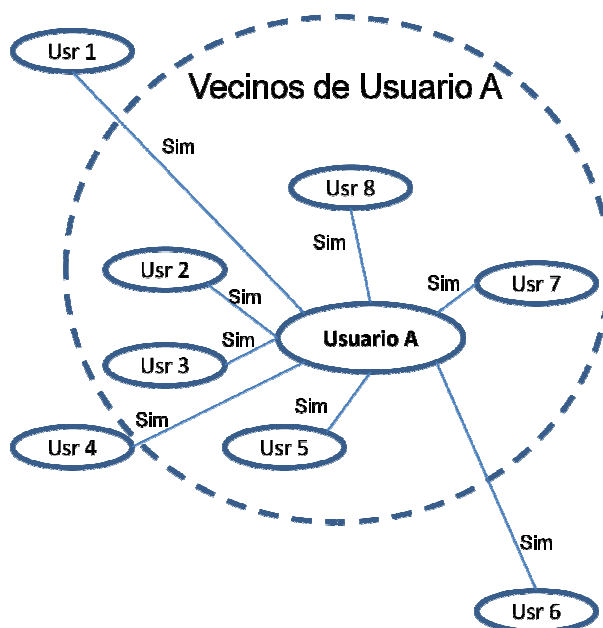


Fig. 43. Cluster de vecinos del usuario A

Una vez que existe un *cluster* de vecinos para cada usuario, se procede a calcular la predicción del ítem deseado para el usuario en cuestión.

De las opciones estudiadas en el estado del arte, se decide la alternativa de los sumatorios de las desviaciones de la media, según (46), para lo que es necesario conocer, para cada uno de los usuarios vecinos de los que exista rating para el ítem en cuestión, ese propio valor de rating y el grado de similitud con el usuario correspondiente (111). Es importante tener en cuenta que esta decisión implica almacenar los valores de rating medio de cada usuario.

(112)

El proceso completo de predicción del rating del ítem B para el usuario A, se muestra en la Fig. 44 siguiendo el ejemplo planteado en la Fig. 43.

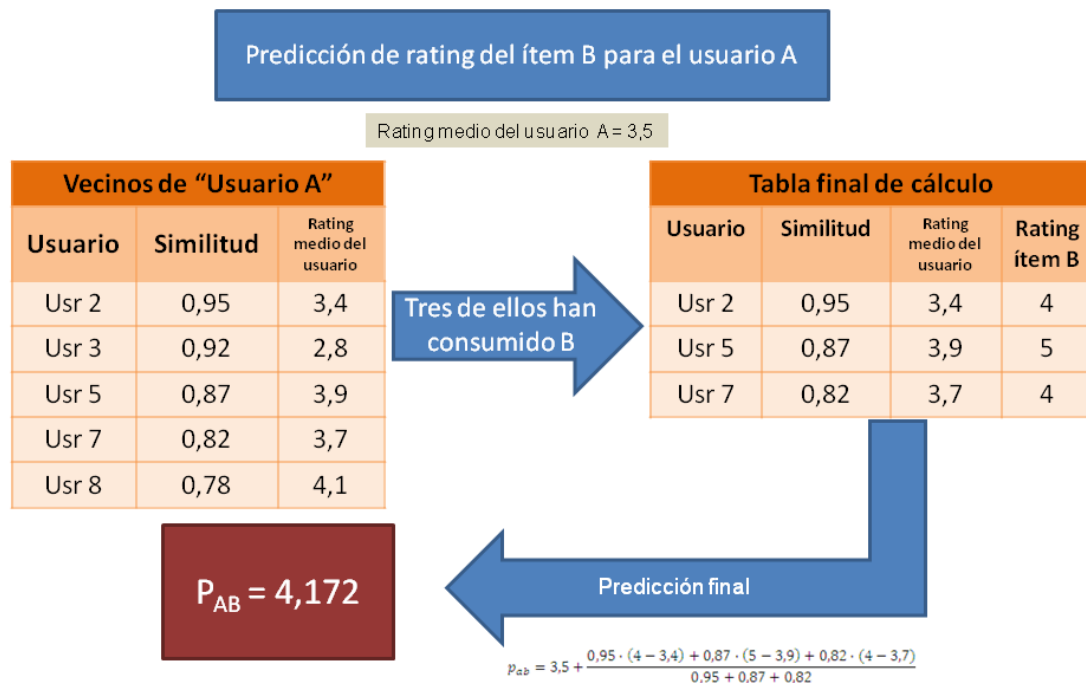


Fig. 44. Ejemplo de predicción del rating del ítem B para el usuario A

6.4 Recomendación de imágenes basada en contenido

Diseño modular

El objetivo del sistema es generar recomendaciones de imágenes de manera personalizada para cada usuario, teniendo en cuenta sus gustos estéticos y sus preferencias por unas determinadas características formales. Las recomendaciones generadas por el sistema están condicionadas, por un lado, a la inferencia de estos gustos por parte del sistema y, por otro, al ajuste de una serie de parámetros de entrada que tienen un valor por defecto pero que el usuario puede modificar para tener el mayor

control posible sobre los algoritmos de recomendación, según los criterios de usabilidad ya definidos en el capítulo de objetivos.

El sistema consta de dos bases de datos y de dos procesos independientes. La primera base de datos, “Anotación de imágenes”, almacena las características generales que se han extraído previamente de las imágenes y que se utilizarán en el proceso de recomendación. La segunda base de datos, “Usuarios”, guarda información de los usuarios, relativa tanto a las imágenes que les gustan y disgustan (ratings explícitos), como a los gustos estéticos personales que han sido previamente inferidos por el sistema. Los dos procesos involucrados en el funcionamiento del sistema, y que interaccionan con estas bases de datos, son los siguientes:

1. **Proceso de clasificación.** Se trata de un proceso *offline*, independiente de los usuarios del sistema. En el momento en que una imagen se añade al corpus general del sistema, este proceso emplea las herramientas y algoritmos de “Clasificación perceptiva”, ya definidos en el capítulo 5 de la tesis, para extraer de las imágenes las correspondientes características y almacenarlas en la base de datos “Anotación de imágenes”.
2. **Proceso de recomendación.** Se trata de un proceso *online*, que se ejecuta cuando el usuario solicita una recomendación, y genera una lista final de recomendaciones personalizadas. Los módulos involucrados en este proceso son los de “Filtrado basado en contenido”, “Filtrado Colaborativo” e “Hibridación”.

El sistema opera directamente con las dos bases de datos implicadas de manera transparente para el usuario, y solo requiere de éste una identificación y un ajuste opcional de los parámetros de entrada del recomendador. Estos parámetros tienen unos valores por defecto que se corresponden con el comportamiento habitual del usuario, pero pueden ser modificados por él en cualquier momento. En el siguiente apartado se detallan tanto los parámetros de ajuste que pueden ser modificados por el usuario, y que por lo tanto conforman la entrada del sistema, como la salida generada por el recomendador.

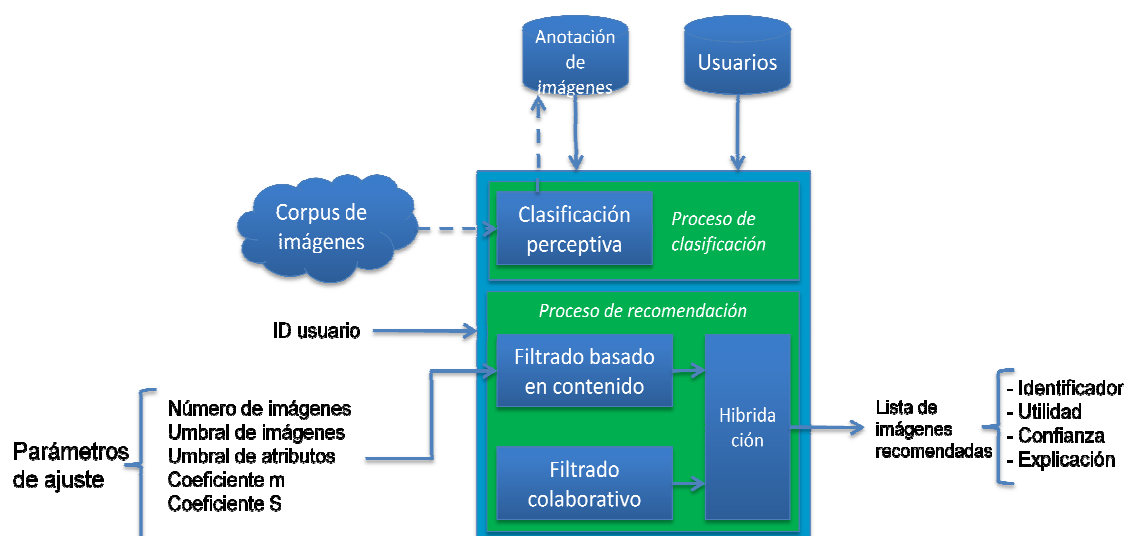


Fig. 45. Esquema de bloques del sistema de recomendación de imágenes

Descripción de entradas y salidas

A continuación se detallan las entradas del sistema, que coinciden con los parámetros de ajuste que el usuario puede modificar cada vez que solicita una recomendación. Se omite, de esta forma, tanto las entradas de las bases de datos, que son transparentes para el usuario, como la propia identificación del usuario.

Entradas:

- **Número de imágenes.** Aplica a la etapa final de hibridación. Especifica el número de imágenes que el usuario quiere que el sistema le recomiende.
- **Peso de las técnicas de filtrado.** Aplica a la etapa final de hibridación. Determina la proporción de influencia que tienen, en la generación de las recomendaciones finales, los dos algoritmos principales del filtrado: el colaborativo y el basado en contenido.
- **Umbral de imágenes.** Aplica al filtrado basado en contenido. Este parámetro es relativo a la exigencia visual de cada usuario; es decir, determina cuál es el umbral para que una imagen guste o deje de gustar al usuario.
- **Umbral de atributo.** Aplica al filtrado basado en contenido. Tiene el propósito de fijar el umbral de impacto de los distintos atributos (características de las imágenes) en el usuario; es decir, determina cuándo un atributo está influyendo en el gusto de ese usuario, ya sea para bien o para mal.
- **Coefficiente m.** Aplica al filtrado basado en contenido. Este coeficiente se introduce en la etapa final del filtrado basado en contenido, aplicándose a las métricas que generan los valores finales de distancia dentro de los espacios vectoriales definidos. De él depende que las recomendaciones finales sean más o menos homogéneas, por lo que su valor debe ir ligado al gusto general del usuario, pero también puede variar según su situación personal, por lo que no debe de tener un valor fijo.
- **Parámetro de suavizado S.** Aplica al filtrado basado en contenido. También depende del usuario, ya que especifica si tienen más influencia en él las características positivas (que le gustan) o las negativas (que no le gustan) de una imagen, además de cuantificar esta diferencia.

Al quedar fuera del campo de investigación que se pretende abarcar, no se han incluido parámetros de ajuste relativos al filtrado colaborativo, aunque queda abierto como posible línea de futuro con el fin de trabajar en las distintas posibilidades de hibridación.

En cuanto a las salidas, el sistema genera una lista de recomendaciones finales, como consecuencia última de la hibridación de los distintos métodos implementados. Cada uno de los elementos de esta lista presenta cuatro atributos que sirven para indicar ciertas características de utilidad para el usuario. Como ya se ha indicado, numerosos estudios han revelado la importancia de parámetros de este tipo en los sistemas de recomendación [144][145].

Salidas:

- **Identificador.** Identifica unívocamente el recurso recomendado, con el fin de que se pueda informar al usuario y que la interfaz pueda mostrarlo.

- **Utilidad.** Indica la predicción del sistema acerca de cuánto va a gustar el recurso al usuario. Procede directamente de conjugar los resultados finales de distancia de las dos técnicas empleadas. En el sistema se ha implementado en una escala porcentual, entre 0 y 100.
- **Confianza.** Muestra cómo de fiable es la recomendación, su grado de certidumbre, ya que esta dependerá de varios factores, como el número de usuarios del sistema, los datos existentes del usuario, la uniformidad de sus gustos, etc.
- **Explicación.** De acuerdo con la predicción de utilidad que se ha realizado, el sistema genera un razonamiento acerca de por qué se recomienda una determinada imagen. El sistema diferencia la utilidad de la predicción procedente de cada una de las dos técnicas y, según este valor, genera explicaciones independientes pero complementarias. Por ejemplo, en el caso de un filtrado colaborativo, una posible explicación sería que se recomienda una imagen debido a que es una de las imágenes preferidas de un tanto por ciento determinado de usuarios afines (aquellos que tienen gustos similares al que pide la recomendación); en el módulo basado en contenido, una posible explicación sería que ciertas características formales de la imagen pueden originar una determinada afinidad con ese recurso. Este parámetro resulta muy valorado en los sistema de recomendación actuales, y es capaz de justificar la validez de los algoritmos [145][146].

Como ya se ha comentado, la principal novedad del sistema radica en los algoritmos de recomendación basada en contenido, por lo que los próximos apartados se centran en explicar el proceso de recomendación de imágenes.

Proceso de recomendación. Introducción

El proceso de recomendación diseñado consta de dos algoritmos principales, que se ejecutan en momentos diferentes dentro del ciclo de vida que mantiene el usuario con respecto al recomendador. El algoritmo principal es el “Algoritmo de recomendación basado en perfil”, que se ejecuta cuando existe información del usuario suficiente como para poder inferir y calibrar sus gustos estéticos y la influencia en sus preferencias de las distintas características formales de las imágenes. Sin embargo, cuando esta información no es suficiente, tiene lugar un problema de *cold start* (relacionado con el número de ítems, ya que el relacionado con el número de usuarios, propio del filtrado colaborativo, ya se solventa con el algoritmo de recomendación basado en contenido). Para evitar este problema se ha definido el “Algoritmo de inicialización basado en contenido”.

A continuación se desgrana el “Algoritmo de recomendación basado en perfil” en dos apartados, uno de ellos referido a la generación de subespacios personalizados y al proceso de clustering, y otro que aborda las métricas definidas para el cálculo de las distancias y el proceso de decisión final. Por último, se explicará brevemente el “Algoritmo de inicialización basado en contenido”.

Para el desarrollo del algoritmo de recomendación se han tenido en cuenta dos métodos clásicos de tratamiento de información, de cuyo estudio se ha partido para desarrollar un nuevo proceso adaptado a las necesidades del sistema. Estos dos métodos clásicos son

el Análisis de Componentes Principales (PCA) [147] y el Escalado Multidimensional (MDS) [148], que no se explican ya que se consideran bien conocidos.

Proceso de recomendación: subespacios personalizados y clustering

La Fig. 46 muestra el proceso completo del “Algoritmo de recomendación basado en perfil”, que está formado por dos ramas principales: la primera de ellas sirve para inferir qué atributos hacen que una imagen resulte atractiva para el usuario, mientras que la rama inferior tiene en cuenta los atributos que hacen que una imagen no sea del gusto del usuario. El objetivo de cada una de las ramas es crear sendos subespacios vectoriales personalizados, el espacio afín y el espacio antiafín. Estos subespacios permiten la representación de las imágenes en sendos sistemas de coordenadas adaptados a los gustos del usuario, sobre los que poder proyectar todas las imágenes que, previamente, tenían una representación vectorial común para todos los usuarios. Estos espacios personalizados permiten la aplicación de métricas y distancias que tienen en cuenta qué características (o combinación lineal de las mismas) influyen realmente en los gustos del usuario, así como la cuantificación de la magnitud en que se produce esa influencia.

A continuación, se explica el proceso para la primera de las dos ramas, ya que el funcionamiento de la otra resulta análogo y es el módulo de decisión final el que combina los resultados a los que se ha llegado por ambos caminos.

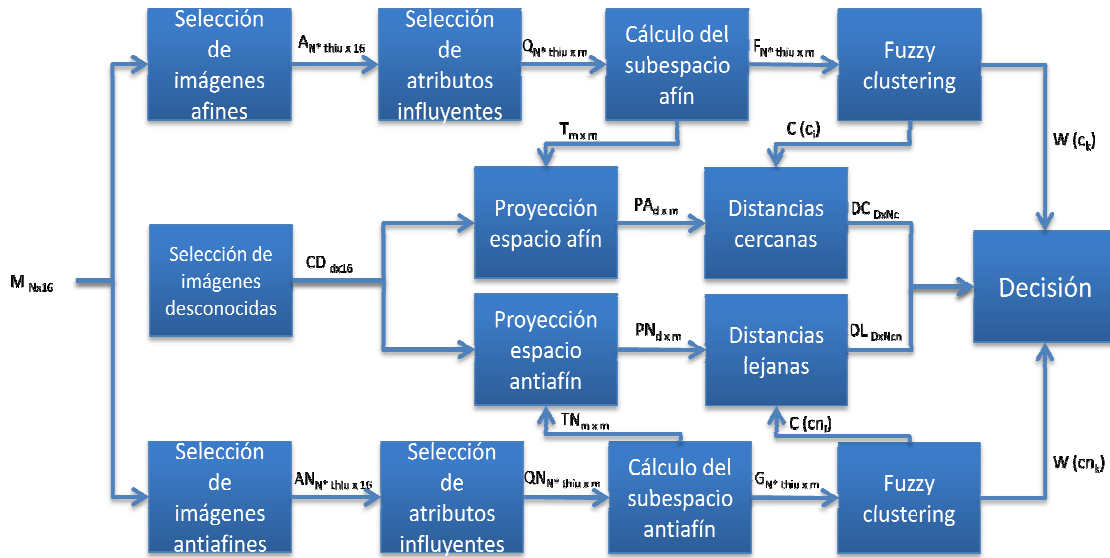


Fig. 46. Algoritmo de recomendación basado en perfil

1. Selección de imágenes afines

Como dato de entrada del sistema se parte del conocimiento de una serie de valoraciones explícitas de un usuario (o medidas de interés inferido) sobre un conjunto de imágenes. Por lo tanto, el primer filtrado necesario establecerá cuáles de esas imágenes son las preferidas del usuario. Teniendo en cuenta que se establece como parámetro de entrada del sistema un umbral thi_u , el sistema considerará para su procesamiento aquellas $N \cdot thi_u$ imágenes que tienen una valoración más alta por parte del usuario. Por lo tanto, la matriz de salida A tendrá unas dimensiones de $N \cdot thi_u \times 16$, que

corresponde al número de imágenes afines (filas) por el número de descriptores generales con el que se caracterizan las imágenes en un primer momento.

La discriminación entre imágenes que gustan e imágenes que no gustan a partir de medidas cuantitativas de preferencia de los usuarios es sólo una de las opciones para manejar estas preferencias. En el recomendador de vídeos se presenta otra alternativa, en la que se ponderan las medidas cuantitativas de preferencia sin necesidad de discriminar entre gusto o no gusto. Esta opción, que también podría ser perfectamente aplicable al recomendador de imágenes, presenta una mayor granularidad, por lo que no se pierde la precisión de la medida del usuario, pero tiene dos inconvenientes: el primero es que disminuye la fiabilidad de la información de entrada, ya que es más difícil discriminar entre un rango amplio de niveles de preferencia que únicamente entre dos; el segundo es que se necesita que las preferencias del usuario estén expresadas de forma cuantitativa en un determinado rango. Por esta razón, en el recomendador de imágenes se prueba el método más flexible a los datos de entrada y se discrimina a partir de las puntuaciones (igual que podría ser, simplemente, a partir de información de gusto o no gusto por unas determinadas imágenes, en cuyo caso esta etapa no sería necesaria) del usuario. El otro caso, de ponderación sobre las puntuaciones, se aplica al recomendador de vídeos, como se explica en el apartado 6.5. De todos casos, ambos métodos serían indistintamente aplicables a recomendadores de imágenes o vídeos, y su elección dependería fundamentalmente de la manera de adquirir los datos de preferencias de los usuarios.

2. Selección de los atributos influyentes

El siguiente paso consiste en decidir qué atributos están afectando realmente a la percepción del usuario. Partiendo de todos los datos de la matriz A , se puede comparar la varianza de cada atributo dentro del grupo de imágenes afines (C_a) con la varianza global de cada atributo dentro del corpus total (C_t). Si la varianza del atributo dentro del grupo de imágenes afines es significativamente inferior que la varianza dentro del corpus (lo cual se regula mediante el parámetro tha_u), el atributo se considera relevante y pasa el segundo filtro. Por lo tanto, se eligen para cada usuario los m atributos que satisfacen la siguiente condición:

$$\frac{\text{Var}(at_i)|_{C_t}}{\text{Var}(at_i)|_{C_a}} > tha_u \quad (113)$$

3. Creación del espacio afín

La relación entre los descriptores influyentes y su grado de relevancia se utiliza para generar cada uno de los vectores del espacio personalizado. Los coeficientes de la matriz T , que indica esta relación, se obtienen como el cociente de las covarianzas dentro del corpus total y dentro del corpus afín para toda combinación de descriptores individuales.

$$T_{ij} = \frac{\text{Cov}(at_i, at_j)|_{C_t}}{\text{Cov}(at_i, at_j)|_{C_a}} \quad (114)$$

La matriz T se utilizará para proyectar cada una de las imágenes definidas en el espacio general de 16 dimensiones sobre el espacio personalizado del usuario. Este método ha sido sugerido por el estudio de los algoritmos ya mencionados de Análisis de Componentes Principales (PCA) y Escalado Multidimensional (MDS).

4. Clustering borroso

Después de la creación del subespacio personalizado, es necesario conocer los puntos de referencia de este espacio sobre los que aplicar las métricas a las imágenes desconocidas (aquellas de las que no tenemos información del usuario acerca de su gusto). Los puntos de referencia se obtienen proyectando, en primer lugar, todas las imágenes afines sobre el nuevo espacio personalizado. Sobre este conjunto de puntos multidimensionales, se aplica un proceso de *clustering* borroso (*Fuzzy Clustering*) para hallar los centroides, que serán considerados puntos de referencia. Cada uno de estos centroides representa un determinado estilo estético de imagen. El *clustering* permite, por lo tanto, considerar que a un usuario le pueden gustar diferentes estilos, y generar las recomendaciones teniendo en cuenta esta consideración. Si solo se considerara un único punto de referencia, éste pertenecería al estilo estético favorito del usuario, y sólo se le recomendarían imágenes de ese tipo, con lo que el recomendador resultaría más previsible y menos útil para el usuario. El proceso de *clustering* aporta diversidad y reduce la sobreespecialización, que es un problema típico de los recomendadores basados en contenido.

Para el desarrollo del proceso, se ha optado por el empleo de un algoritmo de *clustering* borroso ya que, de esta manera, cada imagen puede ser asociada a diferentes *clusters* (estilos) con un peso diferente, algo que es coherente con la naturaleza de las imágenes, que normalmente presentan características propias de distintos estilos. La suma de los pesos de todas las imágenes asociadas a un *cluster* dará, por lo tanto, una idea de la importancia de ese estilo artístico para el usuario. Esto servirá para definir de manera más precisa los gustos del usuario y se utilizará en la etapa final de recomendación.

El algoritmo de *clustering* seleccionado por su simplicidad y por el compromiso entre su eficiencia computacional y los resultados experimentales obtenidos, es el *Fuzzy C-Means* [149]. El principal problema del ajuste de los parámetros de caracterización del algoritmo está en hallar el número total de *clusters* de la partición final. Para ello, se han estudiado múltiples índices de validez [150], y finalmente se ha optado por el “Índice de Xie-Beni”, que ha proporcionado los mejores resultados [151]. Este índice optimiza el ratio entre la variación total de los puntos dentro de cada *cluster* con la separación entre ellos, según la siguiente expresión:

$$XB(c) = \frac{\sum_{i=1}^c \sum_{j=1}^N (\mu_{ij})^m \|x_j - v_i\|^2}{N \cdot \min_{i,j} \|x_j - v_i\|^2} \quad (115)$$

Donde c es el número de clusters, N el número de datos de entrada (número de imágenes en este caso), μ_{ij} el valor de la partición dura, x_j el valor de los datos de entrada y v_i el valor de los centroides obtenidos.

Proceso de recomendación: distancias y decisión

Con el fin de independizar los procesos y reducir coste computacional, los resultados tanto de la generación de espacios estéticos personalizados como del *clustering* borroso se almacenan en la base de datos de perfiles de usuario. Esto es, cada recomendación no implicará la activación de los procesos de creación de espacios y de *clustering*, ya que esto sólo se producirá cuando haya nueva información de *ratings* que procesar (nuevos consumos o nuevas valoraciones explícitas del usuario), y la información podrá recuperarse directamente de la base de datos.

Partiendo, por lo tanto, de esta información de entrada, son necesarios solo tres pasos más para completar una recomendación basada en contenido.

1. Proyección de imágenes desconocidas

Se parte de las d imágenes del corpus desconocidas para el usuario. La proyección de todas estas imágenes genera que la matriz inicial de dimensión $d \times 16$ se transforme en dos matrices adaptadas a los gustos del usuario, una correspondiente al espacio afín (PA) y otra al espacio antiafín (PN).

2. Cálculo de las distancias

Para cada imagen desconocida, se obtienen en paralelo dos grupos de distancias, uno por cada uno de los espacios personalizados. Estas distancias se calculan desde la proyección de la imagen desconocida en el correspondiente subespacio hasta cada uno de los centroides de los *clusters* del subespacio. Por lo tanto, se generan dos matrices de espacios, DC (para el espacio afín) y DL (para el espacio antiafín). Ambas matrices tienen una dimensión $d \times nc$, donde d es el número de imágenes desconocidas del corpus y nc es el número de *clusters*.

La métrica elegida para el cálculo de las distancias, basada en una distancia de Minkowski ponderada similar a la empleada en otros estudios sobre calidad y percepción visual [152][153], se muestra a continuación:

$$d(i, c_k) = \sqrt[p]{\sum_{n=1}^m \alpha_{kn} \cdot [(c_{kn} - \mu_n) - i_n]^p} \quad (116)$$

Donde el coeficiente de peso es:

$$\alpha_{kn} = \frac{\sum_{a=1}^N \frac{c_{kan} - x_{an}}{\sigma_n^2}}{\sum_{a=1}^N \frac{c_{kan} - x_{an}}{\sigma_n^2}} \quad (117)$$

La elección de la distancia de Minkowski sobre las otras opciones analizadas en el estado del arte (además de la razón correspondiente a su uso en otros trabajos similares) corresponde a un razonamiento de máxima generalización posible. De este modo, la distancia más general posible sería la de Mahalanobis, pero hay que tener en cuenta que los datos han sido previamente normalizados y escalados, por lo que no es necesario volver a considerar la distribución estadística de los parámetros. Por lo tanto, la

distancia de Mahalanobis queda reducida a una distancia de Minkowski, para cuya aplicación será necesario encontrar el valor óptimo del parámetro p .

La fórmula expresa el valor de la distancia desde la imagen desconocida i (compuesta a su vez de n dimensiones que son los n atributos proyectados en el espacio afín o antiafín) hasta el cluster c_k , y μ_n representa el valor medio del atributo proyectado n . Después de una serie de pruebas experimentales, se llegó a que el valor óptimo de la constante p está en torno a 2,7. La definición del coeficiente de peso se ha llevado a cabo teniendo en cuenta que aquellos atributos cuyo valor para un usuario están más desviados de la media general de ese atributo (especialmente en atributos con poca dispersión, bajos valores de σ_n), resultarán más relevantes para la caracterización de los gustos del usuario. Para ello, a se define como el número de ítems del corpus y x_a es el valor que tiene el descriptor n del ítem a .

3. Decisión final

La decisión final surge de la combinación de las distancias obtenidas en los dos subespacios personalizados, y tiene en cuenta el peso de los *clusters* para dar más importancia a aquellos estilos que son más del gusto del usuario. Por lo tanto, el rating final de una imagen i generado para un usuario se obtiene de la aplicación de la siguiente expresión:

$$R(i) = \sqrt[m]{\left(\sum_{k=1}^K \frac{w(c_k)}{[d(i, c_k)]^2} \right)^m + \left(\sum_{k'=1}^{K'} \frac{[w'(c_{k'})]^2}{[d'(i, c_{k'})]^{-1}} \right)^m / S} \quad (118)$$

Donde K es el número de *clusters* del espacio afín, K' el número de *clusters* del espacio antiafín, $d(i, c_k)$ la distancia de la imagen i al *cluster* k del espacio afín, $w(c_k)$ el peso del *cluster* k en el espacio afín, $d'(i, c_{k'})$ la distancia de la imagen i al *cluster* k' del espacio antiafín, $w'(c_{k'})$ el peso del *cluster* k' del espacio antiafín, m el coeficiente de Minkowski y S el parámetro de suavizado.

La métrica de decisión final también está basada en la distancia de Minkowski, y tiene en cuenta para el usuario la importancia de los atributos que le influyen tanto para las imágenes que le gustan como para las que le disgustan. El primer término de la raíz m -ésima caracteriza la influencia de los atributos positivos, y genera un rating más alto para imágenes más cercanas a los centroides referenciales que representan los estilos estéticos favoritos del usuario (aquellos que tienen mayor peso). El segundo término contribuye con un rating adicional que es mayor para imágenes más alejadas de los estilos que no gustan al usuario.

S y m son parámetros de entrada para el ajuste del sistema, pero se pueden definir para una configuración automática dependiendo de la realimentación implícita del usuario, modelándolo como un parámetro que busque la mayor satisfacción en las recomendaciones generadas. El parámetro m es el coeficiente de Minkowski, utilizado para ponderar las medidas más extremas, tanto para distancias grandes como para pequeñas. El parámetro S , de suavizado, expresa la diferencia de peso que para un determinado usuario deben tener las características positivas o negativas, ya que hay gente más influida por aquello que le atrae y gente más influida por lo que la repele.

Proceso de recomendación: Algoritmo de inicialización

El algoritmo de filtrado basado en contenido sirve, entre otras cosas, para solucionar el problema que tiene el filtrado colaborativo de *cold start* asociado al número de usuarios en el sistema, ya que la recomendación a un usuario no depende de otros, sino únicamente de las características del contenido. Sin embargo, sigue existiendo un problema de *cold start*, asociado al número de datos necesarios de un usuario para poder caracterizar sus gustos.

Para combatir este problema, se ha creado un algoritmo de inicialización, que genera recomendaciones a partir solamente de la información de una imagen que sea del gusto del usuario. El algoritmo busca, a partir de esta imagen, otras cuya percepción general pueda ser similar, en cuanto a las características de bajo nivel ya comentadas. La información de esas imágenes similares que gustan al usuario se va almacenando hasta que el sistema está en condiciones de lanzar el algoritmo general de recomendación basada en perfil. El algoritmo de inicialización funciona aplicando directamente las métricas de distancia ya comentadas sobre el espacio vectorial genérico de 16 dimensiones que se ha definido para la caracterización de las imágenes.

6.5 Recomendación de vídeos basada en contenido

Fundamento y justificación

El fundamento del algoritmo de recomendación de vídeos basado en contenido es análogo al descrito previamente para las imágenes. Del mismo modo, la idea clave es representar los ítems (vídeos, en este caso) en función de las características que resultan más influyentes para cada usuario, de tal forma que cada recomendación siga unos criterios distintos, adaptados a la percepción y gusto del usuario. Por lo tanto, en este apartado se describen únicamente las diferencias con el algoritmo de recomendación de imágenes.

La distinta naturaleza del corpus empleado condiciona el diseño del algoritmo. Sin embargo, este condicionamiento no está sujeto al tipo de media que se recomienda, sino a otras circunstancias que se detallarán a continuación. Por esta razón, el algoritmo que se ha empleado para la recomendación de vídeos podría emplearse, en un contexto diferente, para la recomendación de imágenes, y viceversa.

En el caso de la recomendación de imágenes, los ratings de los usuarios sobre el corpus se recogieron ex profeso para ejecutar y probar el algoritmo, por lo que la situación era la de un número reducido de usuarios en el que, sin embargo, cada uno de ellos había valorado los 200 ítems del corpus. Es el caso de un escenario controlado. Por otra parte, para la recomendación de imágenes se recurrió a bases de datos externas con valoraciones recogidas de forma abierta de un gran número de usuarios sobre los ítems del corpus. En este caso, por lo tanto, existe un gran número de usuarios, pero cada uno de ellos ha votado un número limitado y desconocido de los ítems del corpus (en este caso, películas). Dado que en el corpus de imágenes se cuenta con un mayor número de información de cada usuario, resulta más fácil y más fiable discriminar qué ítems gustan y qué ítems no gustan al usuario. Además, el usuario se ha visto obligado a valorar

todos los ítems del corpus, por lo que no existe la posibilidad de que haya valorado solo aquello que considera de su gusto o aquello que no lo gusta. En el caso de las valoraciones abiertas empleadas en el recomendador de vídeos, sin embargo, este caso puede producirse, por lo que no resulta fiable realizar una discriminación cualitativa. Por lo tanto, se utilizan las propias valoraciones como “puntos de gusto”, y las diferencias entre ellas son las que marcan las diferencias de preferencia de unos ítems a otros.

Otro ejemplo que obligara a utilizar uno u otro método sería que la información de valoración de los ítems recogida de cada usuario no fuera cuantitativa sino cualitativa. En ese caso, habría que recurrir al algoritmo de recomendación de imágenes (aplicado al tipo de media correspondiente), ya que opera sobre unos criterios iniciales de gusto o disgusto del usuario por un ítem. De esta forma, los dos algoritmos descritos cubren ambas opciones de recogida de información explícita de usuario. En el caso de valoraciones implícitas recogidas a partir del consumo audiovisual, siguiendo el algoritmo propuesto en esta tesis, el caso sería análogo al de valoración cuantitativa.

Esquema general y diferencias con el algoritmo de recomendación de imágenes

El esquema general de recomendación de vídeos basado en contenido sigue el esquema de la Fig. 47.

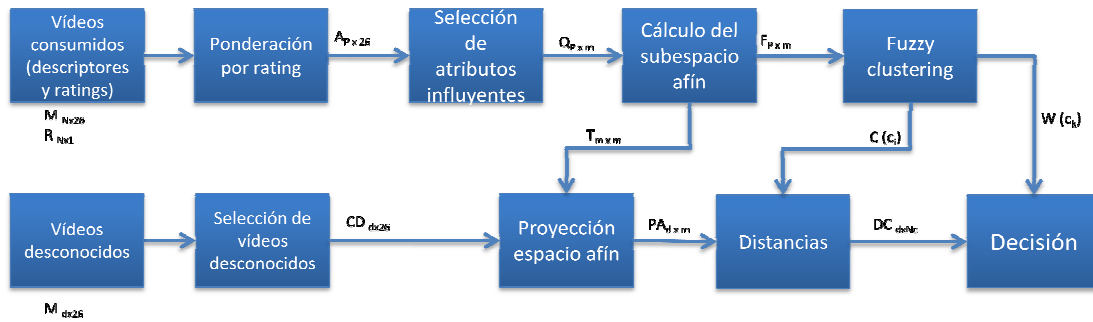


Fig. 47. Proceso de recomendación de vídeos basado en contenido

Como se puede apreciar, la principal diferencia con el esquema de recomendación de imágenes es que desaparece la rama del espacio antiafín y de los ítems que no gustan al usuario. Esto se debe a que todas las preferencias, tanto positivas como negativas, se incorporan a un único subespacio personalizado, subespacio afín, en el que los núcleos de influencia en el usuario están *clusterizados*, y cuya clasificación como *cluster* positivo (representa un estilo que gusta al usuario) o negativo (representa un estilo que no gusta al usuario) se cuantifica según su peso dentro del subespacio. Para ello es necesario introducir algunas variaciones en las fases definidas en la descripción del recomendador de imágenes.

En primer lugar, no se realiza una separación entre elementos afines y antiafines, como en el caso del recomendador de imágenes, aunque esta técnica se podría aplicar igualmente, como ya se ha explicado en el apartado 6.4. En su lugar, se aplica una ponderación diferente a cada elemento según el gusto del usuario. Para ello, se cuenta con una matriz que contiene los 26 descriptores (definidos en el capítulo 4) de los N elementos conocidos (películas valoradas) por el usuario, M_{Nx26} , y otra matriz en la que se establece la preferencia de cada uno de esos N elementos en una escala de valoraciones determinada, R_{Nx1} . Esta matriz de valoraciones sirve para realizar la

ponderación de preferencias a partir de la matriz de descriptores, de tal modo que cada fila, correspondiente a cada uno de los ítems, se replica según el valor de su puntuación correspondiente de la matriz R , de tal manera que se obtiene una matriz $A_{P \times 26}$, donde $P = \sum r_{i1}$, siendo r_{i1} los valores de la matriz R .

A partir de este punto, el proceso es análogo al explicado en el recomendador de imágenes, empleando siempre la matriz de descriptores ponderada en cada una de las etapas. Tal y como se detalla en el capítulo 5, sobre esta matriz se realiza el análisis de la variación de la distribución, para lo cual se realiza el contraste de hipótesis Kolmogorov-Smirnov, en el que se obtiene como resultado la lista de los atributos que influyen en el usuario sobre el cual se está trabajando. En el caso de las imágenes bastaba con un análisis de la varianza debido a la normalidad de las distribuciones de probabilidad de los atributos; en este caso se aplica el análisis Kolmogorov-Smirnov ya explicado, siendo su estadístico asociado el que se utiliza para ponderar la influencia de cada atributo sobre el usuario en cuestión.

Después, se continúa creando el espacio afín y se genera una matriz de transformación a partir de las matrices de covarianzas (114) para, a continuación, aplicar el algoritmo de clustering borroso Fuzzy C-Means sobre la proyección de los ítems conocidos sobre la matriz de transformación. Para fijar el número de *clusters* del proceso, igual que en el caso de las imágenes, se realiza una minimización del índice Xie-Beni (115).

Como resultado de esto se obtiene un grado de pertenencia de cada elemento a cada uno de los clusters. Conviene señalar que, en este caso, los elementos preferidos del usuario están replicados, por lo que incrementan notablemente la densidad de los clusters a los que pertenecen en mayor grado. Por esta razón, para obtener la importancia de cada cluster, que se utiliza posteriormente en el proceso de decisión, se suman los pesos parciales de pertenencia de los elementos replicados. Los *clusters* con mayor peso representan los estilos preferidos del usuario, mientras que los de menor peso son aquellos que le resultan menos afines.

En la fase de predicción y recomendación, los pasos de proyección de los ítems desconocidos y de cálculo de las distancias es análogo al de las imágenes, siguiendo las fórmulas (116) y (117). La diferencia en este último proceso radica en la toma de la decisión de recomendación, ya que no es necesario combinar las medidas de distancias de dos subespacios independientes. La recomendación se produce a partir de los valores de la matriz $DC_{d \times Nc}$, de distancias de cada uno de los ítems desconocidos (d) a cada uno de los clusters del usuario (Nc). Para ello, se genera un rating sobre una escala fija de valores (entre un mínimo m y un máximo M) a partir del valor de la distancia al *cluster* de mayor pertenencia. La estimación del rating R para el ítem i se realiza empleando la siguiente fórmula:

$$R_i = \frac{2}{(w_M - w_m)} \cdot \left(w_c - \frac{w_M + w_m}{2} \right) \cdot \left(\frac{M - m}{d_M - d_m} \right) \cdot \left(d - \frac{d_m + d_M}{2} \right) + \frac{M + m}{2} \quad (119)$$

Donde w_M y w_m son, respectivamente, los valores máximo y mínimo de los pesos de los *clusters* del subespacio del usuario, w_c es el peso del *cluster* al que pertenece el ítem i , M y m son, respectivamente, el máximo y el mínimo de la escala en la que desean establecerse los ratings (por ejemplo, para una escala entre 1 y 5, $m=1$ y $M=5$), d_M y d_m son las distancias máxima y mínima de cualquier ítem a cualquier *cluster* del usuario, y

d es la distancia del ítem i al *cluster* más cercano. De esta forma, el rating final resulta una estimación de la preferencia del usuario por el ítem i en una escala comprendida entre m y M .

Finalmente, el algoritmo de inicialización del recomendador de vídeos es análogo al del recomendador de imágenes descrito en el capítulo 5.

6.6 Explicación

Una vez que se han generado las predicciones de los ítems, ya sean imágenes o vídeos, y se han seleccionado las recomendaciones, es necesario completar un último paso, que se ha tenido en cuenta a lo largo del proceso de creación de los algoritmos de recomendación para cumplir uno de los criterios básicos de usabilidad y transparencia de todo sistema de recomendación [8]: resulta fundamental generar, junto con la recomendación de cada ítem, una explicación acerca de cuál es el motivo por el que se recomienda ese ítem a ese usuario.

Dado que se utilizan varias técnicas para generar la recomendación (filtrado colaborativo y filtrado basado en contenido), cada una de ellas aportará unas razones diferentes, por lo que cada explicación asociada a un ítem tendrá un peso diferente, que corresponde con el peso que cada técnica tiene, en cada caso, en la obtención del rating inferido total.

Filtrado colaborativo

La explicación de una recomendación basada en filtrado colaborativo dada a un usuario u es inmediata, ya que se debe a que el ítem gusta a otros usuarios afines a u . La recomendación de un ítem i a un usuario u puede ser positiva o negativa. El primer caso se considera cuando la predicción de rating es superior a la media de preferencias del usuario u ; si es inferior, se trata de una recomendación negativa.

Si, por ejemplo, la media de valoraciones (como siempre, explícitas o inferidas implícitamente a partir del consumo) de un usuario en una escala entre 1 y 5 es 3.7, un ítem que aparezca en la lista de recomendaciones con una predicción de 4.1 será una recomendación positiva, mientras que una predicción de 3.5 implicaría una recomendación negativa. En el caso de la recomendación positiva, se dan las razones por las que el sistema piensa que el ítem i va a gustar al usuario, es decir, muestra las valoraciones de aquellos usuarios u_n afines a u cuya valoración del ítem i sea superior a la media de cada uno de ellos. En el caso contrario, el sistema muestra las identidades y valoraciones sobre el ítem i de aquellos usuarios a los que no ha gustado el ítem, es decir, cuya valoración es inferior a su correspondiente media de valoraciones.

Con el fin de no abrumar al usuario, la lista de identidades de usuarios influyentes y valoraciones sobre el ítem no se muestra al completo, sino que se acota para aquellos que tienen un mayor peso y, por lo tanto, están por encima de un umbral que el propio usuario debe poder ajustar, según el grado de exhaustividad deseado.

Filtrado basado en contenido

En cuanto al filtrado basado en contenido, la explicación no es tan inmediata, ya que consiste en mostrar qué atributos de los que resultan influyentes para el usuario han sido los determinantes para generar una recomendación. Por lo tanto, primero es necesario detectar cuál es el *cluster* de recomendación asociado al ítem. Esta información se puede obtener contrastando los valores de las matrices DC_{dxNc} y $W(c_k)$ (Fig. 47). Una vez identificado el *cluster*, es necesario hacer la transformada inversa tanto de su centroide asociado como del ítem recomendado, para poder trabajar en un dominio de atributos comprensible por un usuario. Para ello, basta con una sencilla operación matricial:

$$I = \hat{I} \cdot T_u^{-1} \quad (120)$$

$$C = \hat{C} \cdot T_u^{-1} \quad (121)$$

Donde I es la matriz de descriptores del ítem i en el dominio general de atributos, \hat{I} la matriz de descriptores de i en el dominio personalizado y T_u^{-1} la inversa de la matriz de transformación correspondiente al usuario u . Análogamente, C y \hat{C} representan los valores del centroide c en los dominios general y personalizado.

A partir de estos valores obtenidos de I y C , aplicando una distancia Manhattan y ordenando de menor a mayor el resultado para cada descriptor, se obtiene la lista ordenada de influencia de los diferentes atributos a la hora de generar la recomendación. Como en el caso del filtrado colaborativo, el usuario debe tener la opción de elegir la exhaustividad de la explicación, que implica que se muestren más o menos atributos con sus respectivos valores.

Hay que tener en cuenta que los valores de las matrices \hat{I} y \hat{C} ya están ponderados por la influencia de los atributos más influyentes en el usuario, por lo que no es necesario considerar esta influencia al realizar la transformación inversa.

6.7 Hibridación, pruebas y validación

6.7.1 Recomendador de imágenes

En primer lugar, es necesario destacar que, a diferencia del caso de los vídeos, no existían disponibles bases de datos con valoraciones explícitas de usuarios sobre imágenes. Por lo tanto, como se ha indicado en el capítulo 5, la base de datos fue generada a partir de una encuesta en la que se pedía a los usuarios que valoraran en una escala entre 1 y 5, de menor a mayor preferencia, su gusto por las 200 imágenes seleccionadas.

Por lo tanto, para la realización de las pruebas se utilizaron los datos de valoración de las 200 imágenes por parte de los 60 usuarios que colaboraron en la encuesta del Anexo I, ya comentada en el apartado de pruebas del capítulo 4 (4.7). Para cada uno de los usuarios, se suprimieron aleatoriamente de la base de datos con la que interactúa el sistema D valoraciones, entre las 200 totales. Por lo tanto, se considera que el usuario ha valorado un número de $N=200-D$ imágenes, y estas son las empleadas para caracterizar

el sistema y los gustos de los usuarios derivados de ellas. Las valoraciones eliminadas se utilizan posteriormente para validar las predicciones. Es importante considerar que no todos los usuarios tienen los mismos gustos globales: puede haber usuarios a los que gusten muchas pinturas y otros usuarios a los que no guste prácticamente ninguna. Para evitar esta distorsión en las medidas, se tiene en cuenta la media de valoraciones entre todos los usuarios.

Normalmente, el rendimiento de los sistemas de recomendación tiende a ser evaluado mediante el MAE (Mean Absolute Error) [154], pero el sistema no calcula predicciones numéricas de *rating* en la escala de votación, por lo que resulta más representativo estimar cuáles de las *NR* imágenes recomendadas por el sistema están, efectivamente, entre las preferidas de cada usuario.

Para lograr este objetivo, se ha definido la medida “Mejora”. El sistema de recomendación selecciona las *NR* recomendaciones generadas para ese usuario y se calcula la nota media de las puntuaciones reales que el usuario dio en su momento (almacenadas en otra base de datos auxiliar). Entonces, se compara esa media con la media de valoración del usuario al conjunto de imágenes consideradas como desconocidas. Se considera que la “Mejora” es, por lo tanto, el incremento de la media de recomendación por encima de la media del usuario, y está asociada a un número de imágenes recomendadas, *NR*.

$$Mejora_{u,NR} = \frac{\text{Mean}(\text{Ratings}_{\text{recommended_images}})}{\text{Mean}(\text{Ratings}_{\text{unknown_images}})} \bigg|_{u,NR} \quad (122)$$

Para obtener resultados homogéneos, los tests se han desarrollado teniendo en cuenta los parámetros del sistema como valores constantes, excepto el Número de Recomendaciones (NR), que ha variado en algunas pruebas. La Tabla 22 muestra el valor de estos parámetros, que se ha fijado después de una batería de test que ha permitido una optimización de los mismos para el promedio de los usuarios. La personalización adaptativa de estos parámetros, que permitiría obtener mejores resultados, se plantea como una de las posibles líneas futuras de trabajo.

<i>Parámetro de entrada</i>	<i>Valor</i>
Umbral de selección de imágenes	0,15
Umbral de selección de atributos	1,15
Coeficiente m	2,5
Coeficiente S	6
Número de imágenes de entrenamiento	150

Tabla 22. Valores de los parámetros de entrada seleccionados para las pruebas

El primer test de este bloque tiene el objetivo de comparar el rendimiento de las distintas variantes técnicas del sistema de recomendación basado en contenido ya descrito. Cada técnica ha sido probada con 60 usuarios y la mejora final es la media de la mejora para cada uno de ellos. En este test se ha generado una recomendación de cinco imágenes para cada usuario (*NR*=5). La técnica general, que considera las ramas afín y antiafín del algoritmo e incorpora para ambas la ponderación de pesos de los *clusters*, ha obtenido los mejores resultados; la segunda, con resultados muy similares, se diferencia respecto a la primera en que no tiene en cuenta los pesos de los *clusters* de

la rama antiafín; la tercera técnica tiene en cuenta la rama afín y antiafín, pero no aplica a ninguna de ellas la ponderación de pesos de los *clusters*; por último, la cuarta técnica sólo emplea la rama afín del algoritmo, y tampoco le aplica los pesos obtenidos de los *clusters*. La Tabla 23 resume los resultados.

<i>Técnica</i>	<i>Mejora media</i>
Técnica completa	1,1255
Técnica sin pesos para rama antiafín	1,1205
Técnica sin pesos para ninguna de las ramas	1,1195
Técnica que utiliza solamente la rama afín	1,0988

Tabla 23. Comparación de técnicas de recomendación de imágenes para NR=5

El siguiente experimento se realizó para comprobar que un número menor de recomendaciones implica que estas recomendaciones sean mejores para el usuario y, por lo tanto, su rating medio real debería ser mayor.

Como se ha considerado un número de 150 imágenes de entrenamiento, se dispone de hasta 50 imágenes desconocidas para solicitar al sistema. Si el número de recomendaciones solicitadas es 50, entonces la valoración media de las imágenes recomendadas por el sistema será igual que la valoración media de las imágenes desconocidas para ese usuario. Es decir, el ratio será 1. Entonces, si descende el número de recomendaciones solicitadas y el recomendador es eficiente, el ratio resultante debe ir ascendiendo de forma monótona.

La Fig. 48 muestra el resultado del valor medio de mejora para los 60 usuarios conforme se va modificando el número de recomendaciones solicitadas. Las funciones en azul y en rojo representan, respectivamente, las dos primeras técnicas definidas en la Tabla 23, con lo que se puede comprobar una vez más que la ponderación de pesos en la rama antiafín tiene resultados apenas perceptibles para la generalidad de los usuarios.

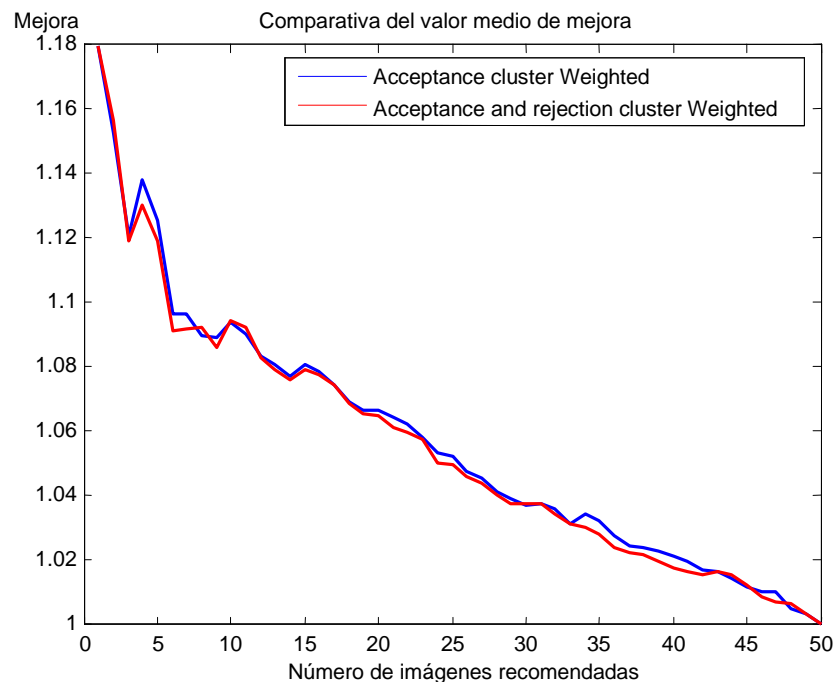


Fig. 48. Valor medio de mejora

Finalmente, se muestran los resultados de la última prueba realizada para el algoritmo basado en contenido en la gráfica de la Fig. 49. La gráfica muestra la mejora que se presenta para cada uno de los 60 usuarios (azul), y se compara con el límite máximo teórico (rojo) al que se podría aspirar si se recomendaran exactamente por orden las imágenes favoritas de cada usuario.

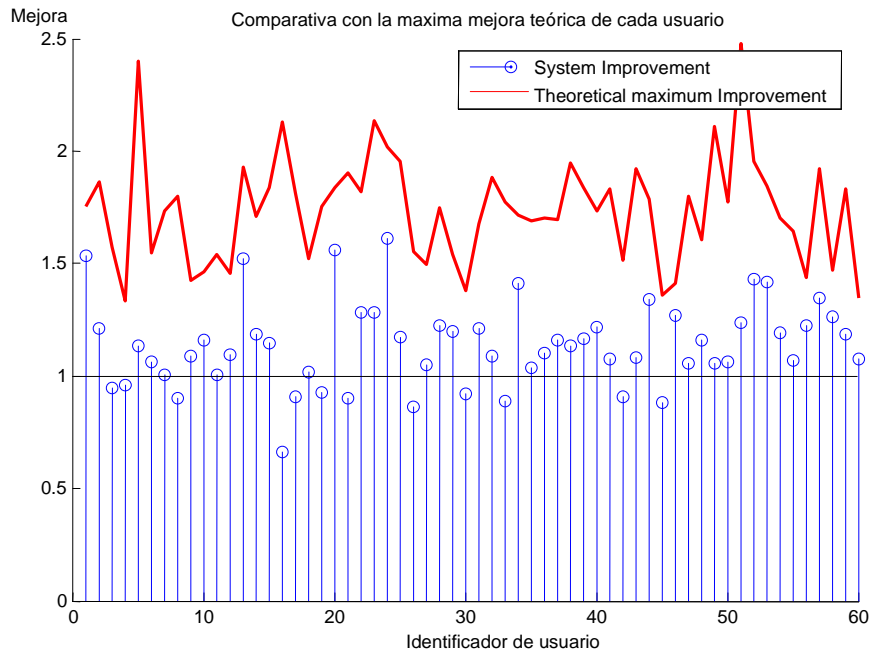


Fig. 49. Comparación con la mejora maxima teórica para $NR=5$

Como se aprecia en la Fig. 49, el sistema de recomendación resulta útil para 48 de los 60 usuarios para el caso en que se solicitan 5 imágenes recomendadas. Después de diferentes pruebas con distintos valores de NR , se detecta que generalmente son los mismos usuarios los que obtienen mejores y peores resultados. Por lo tanto, de cara al sistema híbrido completo, resulta fácil detectar qué usuarios no presentan un buen rendimiento respecto al sistema de recomendación basado en contenido con el fin de aplicar un mayor peso al módulo de filtrado colaborativo. Las técnicas de hibridación inteligente podrían suponer la solución a aquellos usuarios que presentan un comportamiento anómalo con alguna de las técnicas de recomendación desarrolladas.

Ejemplos y resultados cualitativos

Algoritmo de recomendación basado en contenido

De forma ilustrativa del algoritmo de recomendación basado en contenido, se presentan los resultados cualitativos de un ejemplo de ejecución de los algoritmos para un usuario concreto con el corpus de prueba.

Usuario "a"	
Imágenes más recomendadas	Imágenes menos recomendadas
	
	
	
	
	

Tabla 24. Ejemplo recomendación basada en contenido para usuario "a"

Inicialización del algoritmo basado en contenido

Por último, se muestra otro ejemplo concreto de la ejecución del algoritmo de inicialización, partiendo de una de las imágenes del corpus.

Imagen de partida	Imágenes recomendadas
	
	
	
	
	

Tabla 25. Ejemplo inicialización recomendación para usuario “u”

Hibridación con el algoritmo de filtrado colaborativo

Finalmente, se realizan las pruebas de hibridación del algoritmo de recomendación de imágenes. Con ello, además de aumentar la eficiencia de las predicciones, se pretende minimizar el problema de la sobre-especialización, inherente a los recomendadores basados en contenido [144].

Con el sistema híbrido, se han realizado algunos tests sobre el mismo corpus que las pruebas realizadas al recomendador basado en contenido, para proceder a la comparación. Según la definición del apartado 6.3 (en este mismo capítulo), se ha implementado un algoritmo de filtrado colaborativo, y se ha procedido a su hibridación con el algoritmo basado en contenido utilizando los dos métodos más populares, ya descritos en el estado del arte, la hibridación por switching y la hibridación en cascada. También se realizaron pruebas con una hibridación weighted, basada en ponderaciones de las predicciones de los ítems, pero los resultados fueron inferiores.

En primer lugar, se realizaron pruebas únicamente con el algoritmo de filtrado colaborativo, con el fin de obtener un marco comparativo. Estas pruebas, realizadas en las mismas condiciones y de la misma manera descrita para el algoritmo basado en contenido, dieron un resultado de mejora de la media, para listas de recomendaciones de 5 elementos ($NR=5$), de $I_{CF}=1.445$. Este resultado es mejor que el obtenido por el algoritmo basado en contenido antes de realizar ninguna hibridación.

A continuación, se realizaron las pruebas de hibridación.

- La **hibridación por switching** es la más sencilla. Consiste en que el sistema decide, a partir de las medidas de eficiencia en la recomendación calculadas en el pasado, qué técnica de recomendación es mejor para cada usuario. La técnica elegida se aplica a cada usuario para generar las predicciones para los ítems desconocidos. En las pruebas realizadas, el sistema considera que la mitad de las valoraciones del usuario no se han producido y, aplicando las dos técnicas primarias (algoritmo basado en contenido y filtrado colaborativo) a esos ítems, calcula el acierto de cada una de ellas. En función de esos resultados, aplica una u otra técnica para generar la nueva lista de recomendaciones. El resultado de mejora en el mismo escenario es $I_{SW}=1.4678$, que supone solamente una pequeña mejora respecto al filtrado colaborativo.
- En la **hibridación en cascada**, por su parte, el filtrado colaborativo se utiliza para refinar los resultados de los algoritmos de recomendación basados en contenido. La primera etapa (el algoritmo basado en contenido) selecciona aquellas imágenes cuyo estilo y características formales son más afines a un usuario, están más correlacionadas con sus preferencias (se descartan aquellas cuya predicción de valoración está por debajo de la media de valoraciones del usuario). A estas imágenes se les aplica un peso p que se obtiene del estudio de la eficiencia previa del algoritmo. A partir de esta primera selección, el filtrado colaborativo aplica el peso restante $(1-p)$ y genera las predicciones finales de rating. En este caso, la mejora alcanzada es muy significativa, ya que en las mismas condiciones se obtiene un resultado de $I_{HC}=1.66$. Este resultado se acerca al límite teórico de la recomendación, que en este caso tiene un valor de

$I_{TH}=1.8440$. De cualquier manera, la implementación final del sistema debe permitir que el usuario varíe la ponderación de ambas etapas (aunque se le sugiera el valor óptimo), para garantizar la transparencia y configurabilidad del sistema de recomendación que siempre deben estar garantizadas.

La Tabla 26 resume los resultados obtenidos con estas pruebas.

<i>Técnica</i>	<i>Mejora de la media</i>
Recomendación basada en contenido	1.1255
Filtrado colaborativo	1.4450
Sistema híbrido (Switching)	1.4678
Sistema híbrido (Cascada)	1.6600
Máxima mejora teórica	1.8440

Tabla 26. Comparativa general de las técnicas de recomendación de imágenes para NR=5

Análogamente, se probó una técnica en cascada inversa (primero el filtrado colaborativo y después el basado en contenido), y los resultados fueron ligeramente inferiores, lo que tiene su explicación en que el algoritmo basado en contenido es más robusto, mientras que el filtrado colaborativo es más eficiente y, por lo tanto, más útil para el refinamiento sobre un primer filtrado de estilos estéticos afines.

Finalmente, la Fig. 50 compara la mejora de las distintas técnicas de recomendación probadas con el límite teórico para cada uno de los 60 usuarios sobre los que se realizaron las pruebas. El límite teórico varía con cada usuario porque depende de los valores de los ratings que fueron apartados del test inicial para realizar la validación. Este límite considera que, teóricamente, el máximo valor que se puede alcanzar es la media de los ratings de las NR mejores imágenes para cada usuario (de entre las que no ha consumido y, por lo tanto, se consideran para la recomendación), lo cual tiene en cuenta que el objetivo del sistema es recomendar las imágenes favoritas. Por lo tanto, cada mejora sobre un usuario debe ser comparada con su máximo valor teórico.

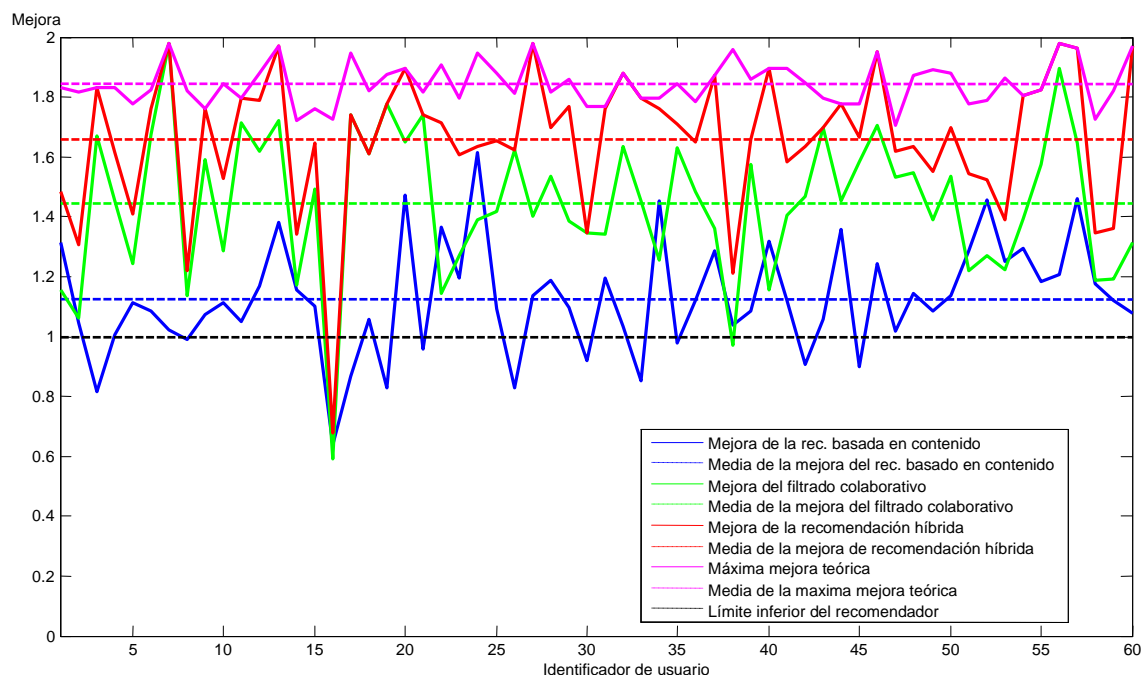


Fig. 50. Comparación entre las técnicas empleadas y la máxima mejora teórica para NR=5

Como se puede apreciar en la Fig. 50, el algoritmo híbrido es útil para 59 de los 60 usuarios, y la recomendación alcanza su máximo teórico para 19 de ellos.

Por último, se ha obtenido el error cuadrático medio (RMSE) de cada técnica (para los 60 usuarios) respecto al máximo teórico, y los resultados se recogen en la Tabla 27.

<i>Técnica</i>	<i>Root Mean Square Error (RMSE)</i>
Recomendación basada en contenido	5.7373
Filtrado colaborativo	3.5626
Recomendación híbrida (Cascada)	2.1380

Tabla 27. RMSE de las distintas técnicas de recomendación de imágenes

6.7.2 Recomendador de vídeos

En el caso del recomendador de vídeos, se dispone de la ya comentada base de datos de Movielens, que presenta 10000054 valoraciones de 71567 usuarios sobre 10681 películas. Sin embargo, como ya se ha señalado, en el corpus de pruebas se debe limitar el número de ítems (películas), ya que es necesario que estas se encuentren previamente anotadas con los descriptores definidos en el capítulo 5 de esta tesis. Por lo tanto, se realiza un primer filtrado de la base de datos empleada, manteniendo la totalidad de usuarios, pero dejando sólo las 200 películas seleccionadas (como se explica en el apartado de pruebas del capítulo 5) y las valoraciones referentes a esos ítems. Por este motivo, la base de datos sobre la que finalmente se trabaja para la realización de las pruebas presenta 952363 valoraciones. Estas valoraciones, junto con los descriptores de los 200 ítems del corpus, componen, respectivamente, la información del usuario y del contenido que utiliza como partida el sistema de recomendación, y que se almacenan en las tablas correspondientes de una base de datos MySQL.

A partir de estos datos de entrada, el sistema ya es apto para un funcionamiento integral. Sin embargo, la validación requiere una comparación de resultados que permita verificar el correcto funcionamiento de los algoritmos, y cuantificar la bondad de sus aproximaciones.

De la misma manera que ocurría en el caso del recomendador de imágenes, es importante recordar que el sistema no tiene como objetivo realizar la predicción más fiable de las valoraciones del usuario, sino generar la recomendación más próxima a la recomendación ideal, es decir, la de aquellos elementos que serán hipotéticamente¹³ más valorados por el usuario. Por esta razón, se descarta emplear el indicador MAE, que únicamente establece el error en la predicción de las valoraciones, y se opta por utilizar nuevamente la medida “Mejora”, definida en (122), a partir de las lista de ítems recomendados, ya que así se aporta información acerca del acierto del recomendador a

¹³ El término hipotéticamente no es gratuito. La validación de un sistema de recomendación debería considerar la utilidad para cada usuario de las recomendaciones que los algoritmos generan, para lo que, en principio, bastaría con comparar el resultado con otras valoraciones desconocidas para el sistema pero conocidas a la hora de realizar la validación. Sin embargo, de esta manera se ignora la propia influencia del motor de recomendación, que puede sugestionar al usuario y modificar lo que sería una valoración obtenida de manera independiente al recomendador. El estudio de estos sesgos es una línea de investigación incipiente en el campo de los sistemas de recomendación.

la hora de recomendar unos ítems en lugar de otros, más allá de la predicción cuantitativa.

Además de este parámetro, se define una nueva medida que permite, para cada usuario, establecer cuánto se ha acercado el sistema a la recomendación perfecta para una solicitud de recomendación de NR ítems. Se denomina “Bondad” a esta medida, y se define para un usuario u y una solicitud de NR recomendaciones:

$$\text{Bondad}_{u,NR} = \frac{\text{Media (Ratings}_{\text{lista_reco_real}})}{\text{Media (Ratings}_{\text{lista_reco_ideal}})} \Big|_{u,NR} \quad (123)$$

La medida “Bondad” se ha introducido porque, al disponer de un número tan elevado de usuarios, no se puede mostrar individualmente cuánto se acerca cada uno de ellos a la recomendación perfecta, tal y como se hacía en el caso de las imágenes en la gráfica de la Fig. 50.

Para aplicar estas dos medidas a los datos disponibles, es necesario hacer un preprocesado de la base de datos de partida, con el fin de ocultar al sistema algunas valoraciones de los usuarios y que estas se puedan utilizar posteriormente para obtener las medidas de acierto. Para ello, se elige aleatoriamente un 50% de las valoraciones almacenadas en la tabla *vratings* y se pasan a una tabla independiente de la base de datos. Esta tabla *vratingsaux*, será empleada únicamente para comprobar el acierto de las recomendaciones, y los ítems con valoraciones para un usuario u formarán el corpus de elementos que el sistema puede recomendar a ese usuario u .

Finalmente, para la realización de las pruebas, se elige únicamente a aquellos usuarios de los que se dispone de un mínimo de 10 películas con valoraciones en la tabla *vratingsaux*, es decir, que se puedan generar recomendaciones sobre un mínimo de 10 películas. Solo 13380 usuarios cumplían esta condición, por lo que este es el número final de usuarios para el que han realizado todas las pruebas y simulaciones.

Para las pruebas se consideran las mismas cuatro técnicas ya probadas para el caso del recomendador de imágenes: recomendación basada en contenido; recomendación basada en filtrado colaborativo; recomendación híbrida por *switching* y recomendación híbrida en cascada. Las dos técnicas básicas, basadas en contenido y en filtrado colaborativo, se han explicado en la parte teórica de este capítulo, y las hibridaciones para los vídeos, análogas al caso de las imágenes, se definen a continuación:

- **Hibridación por *switching*.** El sistema aplica a cada usuario una de las dos técnicas básicas, filtrado basado en contenido o filtrado colaborativo, en función de los resultados de hipotéticas recomendaciones previas. Su implementación se ha realizado de manera análoga al caso de recomendación de imágenes.
- **Hibridación en cascada.** Para la hibridación en cascada, igual que en el caso de las imágenes, primero se filtra la lista de vídeos según la recomendación basada en contenido y, posteriormente, se aplica el filtrado colaborativo para refinar la recomendación. El filtrado basado en contenido, en primer lugar, descarta los ítems cuya predicción está por debajo de la media de valoraciones del usuario. Para los ítems con predicción superior, aplica un peso p , entre 0 y 1, que se

obtiene, como en el caso de la hibridación por *switching*, a partir de la comprobación en el acierto de recomendaciones anteriores (simulando, igualmente, que la mitad de los ítems, seleccionados aleatoriamente, no han sido valorados por el usuario). El filtrado colaborativo, por su parte, aplica un peso $1-p$ a sus predicciones y las suma a las predicciones procedentes del algoritmo basado en contenido, siempre que estas predicciones no hayan sido descartadas en la primera etapa.

La Tabla 28 muestra los resultados de mejora y bondad (métricas definidas en (122) y (123)) de las cuatro técnicas consideradas para un número de recomendaciones solicitadas de $NR=10$.

<i>Técnica</i>	<i>Mejora</i>	<i>Bondad</i>
Recomendación basada en contenido	1.0125	0.9044
Filtrado colaborativo	1.0539	0.9400
Sistema híbrido (Switching)	1.0613	0.9462
Sistema híbrido (Cascada)	1.0648	0.9491
Máxima mejora teórica	1.1185	1

Tabla 28. Comparativa general de las técnicas de recomendación de vídeos para $NR=10$

Se puede comprobar, como ya ocurría en el caso de las imágenes, que la mejora del filtrado basado en contenido por sí solo es muy limitada y queda por debajo de la referencia del filtrado colaborativo. Sin embargo, hibridando sus resultados con los del filtrado colaborativo, los resultados sí son superiores. Dentro de las dos hibridaciones, destaca que la aplicación de la técnica en cascada presenta resultados superiores a los de la hibridación por *switching*, pero la diferencia entre ambas es menor que en el recomendador de imágenes. También es necesario señalar que los resultados de mejora pueden parecer inferiores para los vídeos que para las imágenes, pero en realidad esto no es así, ya que la máxima mejora teórica en el caso de los vídeos es muy inferior al de las imágenes. Esto es debido a la diferente naturaleza de los corpus de datos. En el caso de las imágenes, el corpus estaba muy controlado, ya que se disponía de valoraciones de todos los usuarios sobre todos los elementos; por lo tanto, todos los usuarios tenían un margen de mejora parecido, y contaban con un mismo número de valoraciones de entrenamiento. En el caso del corpus de vídeo, cada usuario u tiene un número Un de valoraciones, con lo que en muchos casos la posibilidad de mejora disminuye debido a la existencia de pocas películas recomendables frente al número NR que se solicitan.

Para completar las pruebas realizadas con $NR=10$, se repiten las simulaciones con un valor $NR=5$, para el cual el margen de mejora, como es lógico, aumenta. La condición de funcionamiento de los recomendadores es que al disminuir el número de recomendaciones NR , la mejora aumente, ya que si se escogen menos ítems, estos deben ser mejores.

La Tabla 29 muestra los resultados de las simulaciones y se aprecia que, en todas las técnicas, se cumple que el resultado de mejora es superior al de las simulaciones realizadas para $NR=10$.

<i>Técnica</i>	<i>Mejora</i>	<i>Bondad</i>
Recomendación basada en contenido	1.0242	0.8521
Filtrado colaborativo	1.0814	0.8995
Sistema híbrido (Switching)	1.1019	0.9168
Sistema híbrido (Cascada)	1.1465	0.9519
Máxima mejora teórica	1.1997	1

Tabla 29. Comparativa general de las técnicas de recomendación de vídeos para NR=5

La Tabla 29 ofrece valores medios de eficiencia, pero no permite calibrar para cuántos usuarios son realmente útiles las distintas técnicas de recomendación. Por esta razón, se muestran unos histogramas en la Fig. 51, junto con sus correspondientes valores numéricos en la Tabla 30, donde se ha obtenido el porcentaje de usuarios que están dentro de los diferentes umbrales de mejora posibles del sistema.

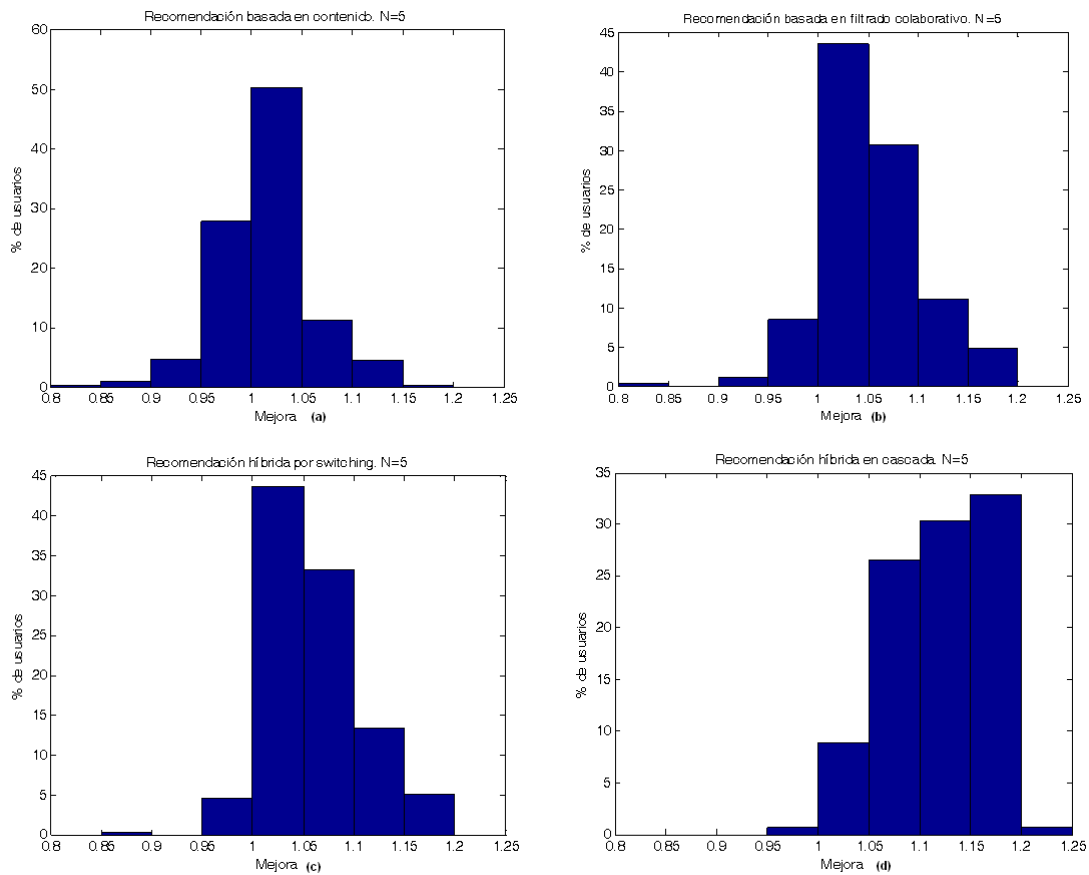


Fig. 51. Comparación de histogramas según las cuatro técnicas de recomendación

El umbral de Mejora=1 resulta clave, ya que es el límite a partir del cual se puede considerar que el recomendador es útil para el usuario, ya que le recomienda contenidos que le gustan por encima de la media. Sobre este dato, es importante ver la cantidad de usuarios para los que cada técnica de recomendación es útil:

- Recomendación basada en contenido: 66.15 % de los usuarios
- Recomendación basada en filtrado colaborativo: 90.19 % de los usuarios
- Recomendación híbrida mediante *switching*: 95.21 % de los usuarios
- Recomendación híbrida en cascada: 99.37 % de los usuarios

Umbrales	Recomendación basada en contenido	Filtrado colaborativo	Switching	Cascada
$U < 0.8$	0	0	0	0
$0.8 < U < 0.85$	0.2604	0.2653	0	0
$0.85 < U < 0.9$	1.0417	0	0.2660	0
$0.9 < U < 0.95$	4.6875	1.0610	0	0
$0.95 < U < 1$	27.8646	8.4881	4.5213	0.6329
$1 < U < 1.05$	50.2604	43.5013	43.6170	8.8608
$1.05 < U < 1.1$	11.1979	30.7692	33.2447	26.5823
$1.1 < U < 1.15$	4.4271	11.1406	13.2979	30.3797
$1.15 < U < 1.2$	0.2604	4.7745	5.0532	32.9114
$1.2 < U$	0	0	0	0.6329

Tabla 30. Porcentaje de usuarios según umbrales de mejora

Ejemplo cualitativo

Algoritmo de recomendación basado en contenido

A continuación se presenta un ejemplo ilustrativo de recomendación basada en contenido sobre el corpus empleado para un determinado usuario y $NR=5$.

Predicción	Título	Atributo más influyente
5.0	Ring (Ringu) (1998)	Complejidad media del movimiento del cuadro
4.99	Talk to Her (Hable con Ella) (2002)	Mediana de la duración de los planos
4.98	A Matter of Life and Death (Stairway to Heaven) (1946)	Intensidad media del movimiento de la cámara
4.97	Notting Hill (1999)	Mediana de la duración de los planos
4.94	Pretty Woman (1990)	Complejidad media del movimiento del cuadro

Tabla 31. Ejemplo de recomendación de vídeo basada en contenido

Inicialización del algoritmo basado en contenido

Por último, se presenta un ejemplo ilustrativo del algoritmo de inicialización del recomendador de vídeos, que en este caso sugiere 5 ítems a partir de un ítem dado.

Ítem de entrada: Unforgiven (1992)		
Predicción	Título	Atributo más influyente
5.0	Seven (a.k.a. Se7en) (1995)	Duración media del plano
4.97	Goldfinger (1964)	Duración media del plano
4.95	The Usual Suspects (1995)	Duración media del plano
4.94	Charade (1963)	Mediana de la duración de los planos
4.94	Million Dollar Baby (2004)	Duración media del plano

Tabla 32. Ejemplo del algoritmo de inicialización del recomendador de vídeo

6.8 Resultados obtenidos

Este capítulo ha presentado los algoritmos de recomendación, es decir, la culminación del trabajo previo, ya que a partir de la información de usuario inferida y la información del contenido extraída, se generan las recomendaciones finales.

Para una validación suficiente, no se han podido utilizar los datos de intereses de usuarios obtenidos a partir del consumo audiovisual inferidos en el capítulo 4 de esta tesis, ya que se necesitaba una cantidad ingente de datos de entrada que no era posible recabar manualmente. Por esta razón, se ha hecho uso de la base de datos *MovieLens 10M*, ya comentada a lo largo del capítulo.

La novedad propuesta por la tesis es el sistema de recomendación basado en contenido basado en características formales y estéticas, que, a pesar de no mejorar por sí mismo otras técnicas ya existentes, sí consigue un gran avance cuando se junta con un algoritmo de filtrado colaborativo mediante técnicas de hibridación, añadiendo, además, novedosos criterios de recomendación de carácter forma.

Según las pruebas realizadas, se ha comprobado que la mejora del algoritmo híbrido creado respecto al filtrado colaborativo para el recomendador de imágenes es de un 11.7%, habiendo obtenido este porcentaje de normalizar y comparar los índices de mejora ya expuestos en el capítulo ($\frac{1.66-1.445}{1.844} = 11.66\%$). En el caso del recomendador de vídeos, este porcentaje se obtiene directamente del resultado de la bondad, y la mejora sobre el filtrado colaborativo es de 5.24%. Esta diferencia entre ambos tipos de recomendadores, como ya se ha señalado, se debe al diferente margen de mejora que llevan implícitos los dos corpus empleados.

Por lo tanto, se puede concluir que la mejora alcanzada se debe a la influencia en los usuarios de ciertas características formales y estéticas de las imágenes y los vídeos. Para una recomendación basada en contenido más completa, sería necesario plantear en el futuro un escenario de recomendación basado en temas y características de tipo semántico, de tal manera que se pudieran fusionar los resultados de ambas aproximaciones. Algunas de las características del audiovisual más importantes de cara a la percepción de los usuarios, como la narrativa audiovisual, están ligadas por igual a la forma y al fondo, por lo que la mencionada unión de ambos tipos de características permitiría la obtención de niveles de abstracción más próximos a la percepción directa e intuitiva de un ser humano.

Por último, para el diseño de los algoritmos se ha tenido en cuenta que estos sean totalmente configurables y transparentes, de tal modo que el usuario pueda modificar el tipo de recomendación que recibe y cambiar la granularidad de la explicación asociada a cada ítem recomendado.

7. Conclusiones y líneas futuras

7.1 Conclusiones

A lo largo de los capítulos centrales de la tesis se ha demostrado el cumplimiento de las hipótesis planteadas en el capítulo de objetivos (capítulo 2), siguiendo para cada una de ellas la metodología de investigación y pruebas que se ha considerado más apropiada, y que también se especifica en ese mismo capítulo.

Es importante reseñar que los algoritmos y técnicas desarrolladas no forman necesariamente un sistema único (aunque puede formarlo), sino que conforman una serie de piezas que se deben articular de la manera adecuada para el caso de uso planteado, tal y como se explica en el capítulo 3 de la tesis. Estas técnicas permiten desarrollar sistemas de recomendación con diferentes requisitos, dependientes del escenario, y con distintas modalidades de uso. Para el desarrollo de estas técnicas se han realizado, fundamentalmente, tres grandes aportes al estado del arte, correspondientes a los tres capítulos centrales de la tesis y a la comprobación de las hipótesis planteadas.

- Se ha trabajado en la inferencia de interés de los usuarios por los contenidos a partir del modelado de su consumo audiovisual. Esta información se puede utilizar directamente para recomendar a través de las preferencias inferidas (ya sea en entornos categorizados o multiatributo), o puede servir como entrada a algoritmos más complejos de recomendación. Los avances principales en esta línea han estado centrado en la mejora de las ponderaciones lineales cuantitativas de tiempo de consumo audiovisual para inferir intereses que todavía se consideraban referencia en el estado del arte. Para ello, además de enriquecer esa ponderación con parámetros cualitativos, se ha creado un modelo probabilístico que tiene en cuenta la redundancia de los distintos contenidos para considerar, de esta forma, las tendencias de consumo personal y la importancia de los factores contextuales.
- Se ha avanzado en el campo de la descripción de contenidos audiovisuales en cuanto a los elementos útiles para realizar la recomendación, centrando los esfuerzos en extraer tanto de imágenes como de vídeos información relevante para definir las preferencias de los usuarios y, por lo tanto, poder recomendar posteriormente de acuerdo con esas conclusiones. Para ello se ha definido una serie de nuevos descriptores visuales con significado estético a partir de características morfológicas y estructurales de imágenes y vídeos, empleando la base teórica suministrada por analistas de la imagen y de la materia fílmica. Por lo tanto, se ha avanzado más allá de las etiquetas semánticas centradas en los temas del contenido que hasta ahora se imponían en los sistemas de recomendación basados en contenido. Para ello, después de la definición, se ha comprobado mediante métodos estadísticos su influencia en un número considerable de usuarios (capítulo 5)

- Por último, se han creado dos nuevos algoritmos de recomendación híbridos de contenidos multimedia, tanto para imágenes como para vídeos. Estos algoritmos incorporan la percepción estética y estructural de los usuarios a las técnicas clásicas de recomendación social como el filtrado colaborativo, mejorando tanto las predicciones como la fiabilidad y la variedad de las recomendaciones. Asimismo, la mejora de las recomendaciones a partir de la toma en consideración de estas características estéticas ratifica la hipótesis ya demostrada en el punto anterior, acerca de la influencia de los descriptores definidos en los gustos de los usuarios.

Además de la comprobación del éxito de las hipótesis, a partir del análisis de las pruebas realizadas se pueden extraer otras conclusiones interesantes sobre las técnicas y sistemas propuestas. Estas se exponen a continuación:

- La aplicación de los modelos de consumo a la recomendación plantea todavía discrepancias relativas al uso de información implícita o explícita, por las diferencias de significación y concepto que acarrearán. En ocasiones, se tiende a considerar que una afirmación explícita del usuario acerca de cuáles son sus preferencias prima sobre cualquier tipo de inferencia que realice una máquina. Sin embargo, esta aproximación no considera que existen factores culturales, sociales o familiares que introducen ciertos sesgos en las opiniones declaradas, de tal modo que esta información explícita puede no ajustarse a la realidad. Este hecho queda patente en las pruebas del capítulo 4 aplicadas a entornos categorizados, en el que el sistema mejora su eficiencia a la hora de predecir los gustos de los usuarios (tomando como referencia la información explícita) conforme aumenta el número de consumos, hasta que aparecen las categorías conflictivas, que se pueden ver influidas por una cierta reputación social. En esos casos, el resultado es inverso: más información para el sistema implica menor conocimiento, lo que, como contradicción, obliga a pensar en la fiabilidad de los datos explícitos: los usuarios no han sido sinceros a la hora de evaluar sus gustos sobre cierto tipo de contenidos. Esto implica que, dependiendo de la aplicación y del tipo de información que deseemos obtener, es necesario estudiar si es preferible emplear información explícita o implícita, ya que esta última puede resultar más fiable en determinados escenarios.
- Los descriptores extraídos de las imágenes y, especialmente, de los vídeos, están pensados con el objetivo de generar recomendaciones; esto es, aunque algunos descriptores podrían utilizarse con otros fines, sería necesario estudiar la validez para su aplicación concreta. Esto se debe a que, en muchas ocasiones, lo que los descriptores pretenden caracterizar es una medida cualitativa aproximada, más adaptada a la percepción del usuario que a una medida absolutamente precisa de la realidad. Un buen ejemplo es el caso de los vectores de movimiento, para cuya extracción se emplean algoritmos de medida del flujo óptico que podrían ser más precisos, pero que presentan una alta eficiencia y son suficientes para el objetivo propuesto. El objetivo de los vectores de movimiento obtenidos es cuantificar el movimiento presente en las imágenes o el movimiento de la cámara tal y como afectan a un espectador, con el objetivo de diferenciar diferentes tipologías fílmicas. Estos métodos seguramente no serían admisibles para realizar operaciones de codificación de vídeo, pero para caracterizar una

percepción humana son más que suficientes, acarreando una carga computacional mucho menos, que, además, permitiría la aplicación de los algoritmos en tiempo real.

- Se ha demostrado la influencia de un conjunto de descriptores en las preferencias de un amplio grupo de usuarios; sin embargo, esto no excluye que otros parámetros que no se han podido extraer no resulten tanto o más influyentes que los obtenidos. La tesis ha buscado el compromiso entre las posibilidades de extracción automática de los descriptores y su potencial influencia en los usuarios. Aun permaneciendo al margen de los descriptores semánticos, en algún caso es posible alcanzar mayores niveles de abstracción sin impedir interoperabilidad entre los distintos tipos de contenidos. Este es el caso de los descriptores relacionados con la composición de la imagen, que deberían dar una idea cualitativa de su armonía visual, la geometría de sus figuras, las simetrías, etc. Después de algunos intentos de extracción, estos descriptores fueron descartados tanto por la complejidad computacional como por la ausencia de resultados suficientemente robustos como para poder caracterizar preferencias relacionadas con este tipo de percepción estética de las imágenes. A pesar de que existen algunas aproximaciones en este sentido en el estado del arte, el análisis de sus resultados denota que estas todavía no se pueden aplicar directamente en un sistema de recomendación.
- La influencia de los descriptores de imágenes y vídeos en los usuarios ha sido comprobada con unos determinados corpus, de imágenes de obras de arte y de vídeos de películas respectivamente. La validación, por lo tanto, está garantizada en estos ámbitos de aplicación, pero sería necesario realizar estudios más exhaustivos para determinar cómo influye el tipo de contenido en este grado de influencia. La dificultad de la validación estriba en encontrar valoraciones públicas de un gran número de usuarios sobre un número suficiente de elementos, por lo que en casos de dominios específicos se hace necesaria la recolección de estas valoraciones mediante técnicas de *crowdsourcing*., es decir, empleando herramientas abiertas en un entorno colaborativo como Internet.
- Una vez obtenidos los resultados de las pruebas de influencia de los descriptores, destaca que los cuatro descriptores más influyentes están relacionados con el montaje del vídeo. Esto resulta relevante tanto para la recomendación de contenidos audiovisuales como para otras posibles aplicaciones que atañan a la personalización, dado que, además, el montaje es una característica que se puede modificar de forma automática.
- Estudiando los resultados de influencia de los descriptores de vídeo, llama la atención que los descriptores relacionados con el audio son los que presentan una influencia más baja sobre los usuarios. Esto se debe, principalmente, a que las técnicas de discriminación de tiempos de habla, silencio, banda sonora y sonidos no están suficientemente maduras. Aunque comprobadas de forma aislada presentan buenos resultados, su aplicación a películas reales ocasiona mayores dificultades, ya que en éstas pueden aparecer simultáneamente diferentes tipos de audio entremezclados, lo cual provoca resultados ambiguos o no suficientemente bien ponderados, que distorsionan el resultado final.

- Los descriptores del bloque “Imagen” del recomendador de vídeos son heredados del recomendador de imágenes. En ambos casos, se ha demostrado que resultan influyentes en el usuario, por lo que se demuestra su aplicabilidad a diferentes dominios.
- El algoritmo de inferencia de preferencias a partir del consumo audiovisual genera una serie de valoraciones que conforman la información del usuario. Esta información, sumada a la que procede de la descripción de los contenidos, se explota en el sistema de recomendación híbrido creado.
- El sistema de recomendación creado basa su novedad en el sistema de recomendación basado en contenido, que se basa, a su vez, en los descriptores previamente definidos. Sobre él, se ha desarrollado una hibridación con un sistema de filtrado colaborativo de referencia, y se ha comprobado que aporta una mejora sustancial de eficiencia.
- La combinación de técnicas en el recomendador híbrido permite la aplicación del sistema a los distintos escenarios presentados, cada uno de ellos con diferentes restricciones funcionales y computacionales, ya que emplea tanto técnicas de tipo social como técnicas basadas en contenido.
- Las técnicas de hibridación probadas han sido obtenidas del estudio del estado del arte. De ellas, la técnica en cascada (aplicando primero el filtrado basado en contenido y posteriormente el filtrado colaborativo) ha ofrecido los mejores resultados, debido a que los dos algoritmos hibridados presentan características complementarias.
 - En el caso del recomendador basado en contenido, se puede afirmar que presenta una alta robustez, aunque una limitada precisión. Esto es, este algoritmo difícilmente va a recomendar elementos que no gusten al usuario; sin embargo, no siempre recomienda los elementos que más gustan.
 - El recomendador social, sin embargo, teniendo una mayor precisión, es menos fiable, ya que con frecuencia realiza predicciones muy favorables para elementos que disgustan al usuario.

Esta complementariedad provoca que la hibridación en cascada sea la más adecuada. Además, tras las pruebas realizadas utilizando un método en cascada inverso (aplicando primero un filtrado colaborativo y después un filtrado basado en contenido), los resultados fueron inferiores, lo que resulta coherente con el estudio de las características de ambos algoritmos. Dado que el filtrado basado en contenido aporta mayor seguridad, es lógico que deba ser aplicado en primer lugar, ya que elimina aquellos elementos que menos gustan al usuario; por su parte, el filtrado colaborativo ya no tiene posibilidad de error, al operar sobre una base de elementos seguros, que, al menos, no disgustan al usuario.

- La aplicación del sistema de recomendación híbrido al corpus de prueba utilizado consigue unos resultados muy cercanos al punto crítico, ya que suponen una aproximación del 95% al resultado óptimo, lo cual es necesario considerar de cara a plantear las líneas futuras de investigación.

7.2 Avance del conocimiento y divulgación

El avance en el estado del arte llevado a cabo en esta tesis, resumido en las conclusiones, se ha plasmado en una serie de publicaciones científicas internacionales, que se indican en la siguiente tabla:

Contenido de la tesis	Publicaciones
Capítulo 3 (Técnicas y arquitecturas de sistemas de recomendación multimedia)	<ul style="list-style-type: none"> • M. Alduan, F. Sanchez, F. Alvarez, D. Jimenez, J.M. Menéndez, C. Cebrecos, "System architecture for enriched semantic personalized media search and retrieval in the future media internet". IEEE Communications Magazine, ISSN: 0163-6804, Vol 49 Issue:3, pags. 144 - 151, March 2011. (JCR Impact Factor: 3.785)
Capítulo 4 (Modelado del consumo audiovisual del usuario para la obtención del interés)	<ul style="list-style-type: none"> • Sanchez, F.; Alduan, M.; Alvarez, F.; Menendez, J.; Baez, O.; "Recommender System for Sport Videos based on User Audiovisual Consumption," <i>IEEE Transactions on Multimedia</i>,. doi: 10.1109/TMM.2012.2217121. Vol 14, issue 6 (JCR Impact Factor: 1.935) (Aceptado para publicación, fecha esperada diciembre 2012) • F. Sanchez, M. Barrilero, F. Alvarez, G. Cisneros, "Embedded audiovisual recommender system for user terminals based on user social and implicit information". IEEE International Conference on Consumer Electronics, ISSN: 2158-3994, pag. 773, Las Vegas (USA), January 2011.
Capítulo 5 (Análisis y clasificación de contenidos audiovisuales para la recomendación) y capítulo 6 (Algoritmos de recomendación híbrida)	<ul style="list-style-type: none"> • Faustino Sanchez, Marta Barrilero, Silvia Uribe, Federico Alvarez and Agustin Tena, José Manuel Menéndez. "Social and Content Hybrid Image Recommender System for Mobile Social Networks". <i>Mobile Networks and Applications</i>, Online first. August 2012 (JCR Impact Factor: 0.838) (Aceptado para publicación, fecha esperada diciembre 2012). • M.Barrilero, S. Uribe, M. Alduan, F. Sánchez, F. Álvarez, "In-network content based image recommendation system for Content-aware Networks", IEEE INFOCOM 2011 International Workshop on Future Media Networks and IP-based TV, Shanghai, April 10-15, 2011.

Tabla 33. Difusión de las investigaciones centrales de la tesis doctoral

Además, se han realizado otras publicaciones en congresos internacionales, no relacionadas con el contenido central de la tesis sino con algunos asuntos colaterales, de implementación, de infraestructura de pruebas, o de ideas embrionarias para los desarrollos posteriores:

- F. Sánchez, A. Cerezo, L. García and F. Álvarez, "Interconnected platform for testing services, applications and products based on the DVB-MHP standard".

IEEE International Symposium on Broadband multimedia systems and broadcasting, Bilbao, May 2009.

- F. Sánchez, A. Cerezo, M.L. García, F. Álvarez and J.M. Menéndez, “Remote access platform for applications and T-government services based on the DVB-MHP standard”, 13th IEEE International Symposium on Consumer Electronics, Kioto, 2009.
- D. Abadía, J. J. Navamuel, J. Vea-Murguía, F. Álvarez, J. M. Menéndez, F. A. Sánchez, G. Fernández, M. Rovira and A. Domech, “i-LAB: Distributed Laboratory Network for Interactive TV Set-Top Box and Services testing”, 13th IEEE International Symposium on Consumer Electronics, Kioto, 2009.
- M. Alvargonzález, L. Arnáiz, F.A. Sánchez, and F. Álvarez, "Personalization of Media Delivery in Seamless Content Delivery Networks". 1st International ICST Conference on User Centric Media- UCMedia 2009, ISBN: 978-3-642-12629-1, 2009.

También se han realizado diversas publicaciones divulgativas, entre las que destacan, por su relación directa con el contenido de la tesis:

- Alduán M., Arnaiz L., Cerezo A., Sánchez F., Uribe S. “Tecnologías de personalización de contenidos multimedia”. White paper. Buscamedia, hacia una adaptación semántica de medios digitales multirred-multiterminal. 14 de octubre de 2011
- Cine y publicidad. Soñando las ambigüedades de un futuro tecnológico. Shangrila derivas y ficciones aparte nº 16. La imagen absoluta. Cine y publicidad. Pags. 129 – 142. Octubre 2012. ISSN: 2172-0363

Por último, se han publicado más de 20 textos y artículos de análisis fílmico en revistas nacionales especializadas, que han servido para desarrollar algunas de las ideas sobre las que se construye el capítulo 5 de la tesis (Análisis y clasificación de contenidos audiovisuales para la recomendación). Entre ellos, destaca la publicación de un libro especializado en el análisis de dos textos audiovisuales:

- Aarón Rodríguez y Faustino Sánchez. Retratos de familia – Tránsitos del cine. Infiel (I. Bergman – L. Ullman)/Yi Yi (E. Yang). Shangrila Textos aparte. Colección Intertextos 1. ISBN: 978-84-939366-0-0

7.3 Líneas futuras

El trabajo de investigación desarrollado en esta tesis continúa varias líneas de investigación que ya existían en el estado del arte, además de abrir otras nuevas. De estas últimas, las más destacadas son las siguientes:

- Aplicación de técnicas semánticas (como por ejemplo, la utilización de ontologías, taxonomías o bases de datos sobre contenidos, preferencias...) para:
 - Enriquecer el conocimiento del usuario a través del análisis de su actividad social (comentarios o diferentes acciones sobre el vídeo). Esta línea está relacionada con la creciente convergencia entre sistemas de distribución de contenidos y redes sociales, existiendo ya plataformas que combinan ambas funcionalidades. Esto, al introducir información complementaria, permitiría aumentar la fiabilidad de la inferencia del interés. Por ejemplo, un comentario positivo que un usuario introdujera acerca de un vídeo después de consumirlo aumentaría la robustez del sistema.
 - Enriquecer la información de consumo del usuario con la información explícita suministrada y con otros tipos de información implícita inferida a partir del seguimiento de intereses. Esto permitiría, además de obtener mayor información del usuario (que implica mayor fiabilidad en la recomendación), combatir el *cold start*, ya que la dependencia de los primeros consumos audiovisuales sería menor, debido a que esta carencia se podría compensar con información adicional, adecuadamente tratada, procedente de fuentes externas.
- Aplicación de técnicas automáticas de identificación de personas (usuarios e invitados) para reducir la incertidumbre de la información de entrada al módulo “Factor de compañía”, cuyo éxito, tal y como está definido en esta tesis, depende de la voluntad del usuario de identificar a sus acompañantes. Esto evitaría la pérdida de datos observada en algunas de las pruebas con usuarios reales desarrolladas en esta tesis, debido a que estos usuarios, en ocasiones, olvidaban identificar ante el dispositivo a todos los usuarios e invitados presentes en el momento del consumo.
- Detección de condiciones de consumo anómalas asociadas a la tendencia, y mejora del ajuste de la subjetividad. Por ejemplo, el consumo continuado de contenidos del gusto del usuario puede provocar una tendencia a la baja en cuanto a la manifestación explícita del interés de los siguientes contenidos que se consumirán. Tener en cuenta esta característica podría aumentar la fiabilidad de las predicciones y, sobre todo, conseguir que el sistema fuera más flexible ante cambios en el contexto del usuario. En cualquier caso, esta característica debería incorporarse dentro de las medidas de validez y precisión, y nunca condicionar directamente las predicciones.
- Mejora de la eficiencia de las técnicas de detección de habla, silencio y otros sonidos aplicados a un contenido. El limitado rendimiento actual de estas técnicas aplicadas a entornos complejos (como el de una película) provoca que no se haya podido demostrar la gran influencia que pueden llegar a tener estas características en una gran mayoría de usuarios. La mejora de estas técnicas, además, permitiría obtener descriptores derivados de estos más elaborados.

- Aplicación de técnicas de procesamiento del lenguaje natural y ontologías al audio de los vídeos para establecer los temas tratados y, además, poder detectar escenas con mayor fiabilidad y se capaz de descubrir secuencias, de significado semántico pleno. Esto también podría realizarse a partir de archivos de subtítulos, en el caso en que las técnicas de transcripción no fueran suficientemente fiables o que hubiera una disponibilidad inmediata de estos archivos de subtítulos.
- Adquisición y modelado de características propias de un contenido audiovisual, como información del autor, fecha, actores o personajes presentes, etc. Esta información, disponible en varias bases de datos de carácter público y privado, puede ser utilizada para completar la información de un contenido desde puntos de vista de más alto nivel, generando datos que se pueden explotar con facilidad mediante diversas técnicas de recomendación.
- Hibridación de los algoritmos de recomendación basados en contenido (además de las técnicas sociales) con técnicas semánticas como el *matchmaking* [137] o el análisis formal de conceptos. En este caso, sería interesante contar con información de más alto nivel, como la que se ha propuesto en el punto anterior, para complementar la información de carácter estético ya propuesta en esta tesis.
- Introducción de factores contextuales en el algoritmo de recomendación, que tengan en cuenta las diferentes situaciones en que se encuentre un usuario, y su relación con las condiciones espaciotemporales, el dispositivo, o el tipo de red de acceso (además de otras características de más alto nivel como el estado de ánimo). De esta forma, la recomendación será más flexible y podría permitir, además de mayores posibilidades de configuración y ajuste, mayores capacidades de personalización.
- Introducción de sesgos temporales a la hora de explotar las preferencias de los usuarios, de tal forma que una preferencia inferida recientemente tenga mayor peso que preferencias inferidas (o manifestadas por el usuario) en el pasado. Esta línea, abierta recientemente por el recomendador de Netflix [155], ha empezado a dar muy buenos resultados.
- Ampliación del corpus de contenidos. Si bien las pruebas realizadas en la tesis cuentan con un número más que suficiente de usuarios, el número de elementos disponibles para la recomendación puede ser mucho mayor. Esto implica realizar previamente una anotación masiva de contenidos, siendo para ello necesario, además, conseguir esos contenidos (que deben ser de conocimiento público para poder contar con valoraciones de bases de datos existentes).
- Aplicación de las conclusiones derivadas de la influencia de descriptores a sistemas relacionados, como los que abordan la generación automática de contenidos, o el resumen automático de vídeos. Como ya se ha visto en las conclusiones, los descriptores de montaje son los más influyentes en los usuarios, y esta es una característica que se puede aplicar a cada usuario a la hora de generar automáticamente un contenido audiovisual adaptado a sus preferencias.

8. Referencias

- [1] Sung Y. J.; Jeong-Hee H.; Taek-Soo K.; "A statistical model for user preference" *IEEE Transactions on Knowledge and Data Engineering*, vol.17, no.6, pp. 834-843, Junio 2005
- [2] El-Sherif, S.M.; Eberlein, A.; Far, B.; "Calculating the strength of ties of a social network in a semantic search system using hidden Markov models" *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 2755-2760, 9-12 Octubre 2011
- [3] Lui, B.; Chiu, D.K.W.; Haiyang Hu; Hua Hu; Yi Zhuang; "Ontology Based Content Management for Digital Television Services" *IEEE International Conference on e-Business Engineering, 2009 (ICEBE '09)*, pp.565-570, 21-23 Octubre 2009
- [4] Zhi-Sheng C.; Jang, J.-S.R.; Chin-Hui L.; "A Kernel Framework for Content-Based Artist Recommendation System in Music" *IEEE Transactions on Multimedia*, vol.13, no.6, pp.1371-1380, Diciembre 2011
- [5] Reed, J.; Chin-Hui Lee; "Preference Music Ratings Prediction Using Tokenization and Minimum Classification Error Training," *IEEE Transactions on Audio, Speech, and Language Processing*, vol.19, no.8, pp.2294-2303, Noviembre 2011
- [6] Bo Shao; Ogihara, M.; Dingding Wang; Tao Li; , "Music Recommendation Based on Acoustic Features and User Access Patterns," *IEEE Transactions on Audio, Speech, and Language Processing*, vol.17, no.8, pp.1602-1611, Noviembre 2009
- [7] Adomavicius, G.; Tuzhilin, A.; "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering*, vol.17, no.6, pp. 734- 749, Junio 2005
- [8] Symeonidis, P.; Nanopoulos, A.; Manolopoulos, Y.; "Providing Justifications in Recommender Systems," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol.38, no.6, pp.1262-1272, Noviembre 2008
- [9] Sanchez, F.; Barrilero, M.; Alvarez, F.; Cisneros, G.; "Embedded audiovisual recommender system for user terminals based on user social and implicit

information” *IEEE International Conference on Consumer Electronics*, pp. 773-774, Enero 2011.

- [10] Alduan, M.; Sanchez, F.; Alvarez, F.; Jimenez, D.; Menéndez, J.M.; Cebrecos, C.; "System architecture for enriched semantic personalized media search and retrieval in the future media internet". *IEEE Communications Magazine*, vol 49, no.3, pp. 144 - 151, Marzo 2011
- [11] Estándar MPEG-7, parte 3. ISO/IEC 15938-3:2002. Information technology -- Multimedia content description interface -- Part 3: Visual
- [12] Sanchez, F.; Alduan, M.; Alvarez, F.; Menendez, J.; Baez, O.; "Recommender System for Sport Videos based on User Audiovisual Consumption," *IEEE Transactions on Multimedia*, vol 14, no.6 (Aceptado para publicación, prevista en diciembre 2012)
- [13] Burke, R.; Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction*. Springer, pp. 331-370, vol 12, no4. Noviembre 2002.
- [14] Hyoseop S.; Minsoo L.; Eun Yi K.; “Personalized Digital TV Content Recommendation with Integration of User Behavior Profiling and Multimodal Content Rating”. *IEEE Transactions on Consumer Electronics*, pp.1417-1423, vol.55,no.3, Abril 2009.
- [15] Zhiwen Y.; Xingshe Z.; Liang Z.; Kejun D.; “A hybrid similarity measure of contents for TV personalization” *Multimedia Systems*, pp. 231-241, vol.5, no.4-5, Mayo 2010.
- [16] Chao Liang; Changsheng Xu; Hanqing Lu; “Personalized Sports Video Customization Using Content and Context Analysis”, Hindawi Publishing Corporation. *International Journal of Digital Multimedia Broadcasting*, pp. 1-20, vol. 2010.
- [17] Syeda-Mahmood T.; Ponceleon D.; “Learning Video Browsing Behavior and Its Application in the Generation of Video Previews”; *Proceedings of the 9th ACM international conference on Multimedia*, pp.119-128, Ottawa, Canada. 2001.
- [18] Zimmerman J.; Kurapati K.; Buczak L.; Schaffer D.; Gutta S.; Martino J.; Chapter 5. “Design of a TV Show Recommender Engine and Interface”. TV Personalization System, Springer, pp. 27-51, 2004.
- [19] Billsus D.; Pazzani M.J.; “Learning Collaborative Information Filters”, *Fifteenth International Conference on Machine Learning*, Wisconsin, USA, pp. 46-54, 1998

- [20] Duda R.; Hart P.; "Pattern Recognition and Scene Analysis". New York: John Wiley & Sons, 1973
- [21] J. R. Quinlan: 1983, "Learning Efficient Classification Procedures and their Application to Chess End Games". In R. S. Michalski, J. G. Carbonell and T. M. Mitchell (eds.): Machine Learning: An Artificial Approach, Vol. 1. Palo Alto, California: Morgan Kaufmann Publishers Inc.
- [22] Álvarez F.; Martín C.A.; Alliez D.; Tonda Roc P.; Steckel P.; Menéndez J.M.; Cisneros G.; Jones S.T.; "Audience Measurement Modeling for Convergent Broadcasting and IPTV Networks", *IEEE Transactions on Broadcasting*. Junio 2009.
- [23] TV-Anytime phase 1: S-3 1.1 Metadata - Corrigenda 2 to S-3 V1.1 COR2_SP003v11.zip (15-8-2003) available at <ftp://ftp.bbc.co.uk/Specifications>
- [24] Piesing, J.; "The DVB Multimedia Home Platform (MHP)", *The Institution of Electrical Engineers*, IEEE, Savoy Place, London, 1999
- [25] Merkel, K.; , "Hybrid broadcast broadband TV, the new way to a comprehensive TV experience," *14th ITG Conference on Electronic Media Technology (CEMT)*, vol., no., pp.1-4, 23-24 March 2011
- [26] Sanchez, F.; Cerezo, A.; Garcia, L.; Alvarez, F. "Interconnected platform for testing services, applications and products based on the DVB-MHP standard". *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, 2009.
- [27] Sanchez, F.; Cerezo, A.; Garcia, L.; Alvarez, F.; Menendez, J.M. "Remote access platform for applications and T-government services based on the DVB-MHP standard". *The 13th IEEE International Symposium on Consumer Electronics (ISCE2009)*. 2009.
- [28] www.gatv.ssr.upm.es/~fsg/TVconsumption (Consultado el 20/09/2012)
- [29] Norman M. Bradburn, Seymour Sudman, Brian Wansink. "Asking Questions: The Definitive Guide to Questionnaire Design - For Market Research, Political Polls, and Social and Health Questionnaires", Revised Edition. ISBN: 978-0-7879-7088-8. 448 pages. April 2004
- [30] Rodgers J.L.; Nicewander W.A.;. "Thirteen ways to look at the correlation coefficient". *The American Statistician*, vol. 42, no1, pp. 59–66, Febrero 1988
- [31] Johnston, I. "I'll give you a definite maybe: An introductory handbook on probability, statistics, and excel [Online]", Mayo 2000. Available: www.mala.bc.ca/~johnstoi/maybe/maybe4.htm (Consultado el 20/09/2012)

- [32] Martinez, A.; Arias, J.; Vilas, A.; Garcia Duque, J.; Lopez Nores, M.; , "What's on TV tonight? An efficient and effective personalized recommender system of TV programs," *IEEE Transactions on Consumer Electronics*, , vol.55, no.1, pp.286-294, February 2009
- [33] Xiao Wu; Yongdong Zhang; Junbo Guo; Jintao Li; , "Web video recommendation and long tail discovering," *IEEE International Conference on Multimedia and Expo*, vol., no., pp.369-372, June 23 2008-April 26 2008.
- [34] H. Shin, M. Lee, E. Y. Kim. "Personalized Digital TV Content Recommendation with Integration of User Behavior Profiling and Multimodal Content Rating". *IEEE Transactions on Consumer Electronics*, vol. 55, no 3, pp. 1417-1423. April 2009.
- [35] Sotelo, R.; Blanco-Fernandez, Y.; Lopez-Nores, M.; Gil-Solla, A.; Pazos-arias, J.; , "TV program recommendation for groups based on multidimensional TV-anytime classifications," *IEEE Transactions on Consumer Electronics*, vol.55, no.1, pp.248-256, February 2009
- [36] M.Barrilero, S. Uribe, M. Alduan, F. Sánchez, F. Álvarez, "In-network content based image recommendation system for Content-aware Networks", *IEEE INFOCOM 2011 International Workshop on Future Media Networks and IP-based TV*, Shanghai, April 10-15, 2011
- [37] Faustino Sanchez, Marta Barrilero, Silvia Uribe, Federico Alvarez and Agustin Tena, José Manuel Menéndez. "Social and Content Hybrid Image Recommender System for Mobile Social Networks". *Mobile Networks and Applications*, Online first. August 2012 (Aceptado para publicación, fecha esperada: diciembre 2012).
- [38] T. Adorno. *Teoría estética* (Ediciones Akal, 2005). ISBN: 978-84-460-1670-0
- [39] Gombrich E. H. *Art and Illusion: A Study in the Psychology of Pictorial Representation*. London: Phaidon Press, 2002.
- [40] Aumont, J. *L'image*, París, Mathan, col. <<Faccinéma>>. Las principales teorías de la imagen. Trad. Cast.: La imagen, Barcelona, Paidós, 1992 ISBN: 978-84-7509-744-2.
- [41] Aumont, J., Bergala A., Marie M., Vernet M. *Estética del cine: espacio fílmico, montaje, narración, lenguaje*. Traducción de Nuria Vidal. Ediciones Paidós Ibérica, 2005, Argentina y Uruguay. ISBN: 978-950-12-7517-9
- [42] Itten, Johanness. *The Art of Color: The Subjective Experience and Objective Rationale of Color*. New York: John Wiley, c2002. ISBN: 0471289280.

- [43] Davis, Steven. *Color perception: Philosophical, Psychological, Artistic, and Computational Perspectives*. New York: Oxford University Press, 2000. ISBN: 0195136675.
- [44] Burch, N. *La lucarne de l'infini*. Naissance du langage cinématographique, París, Nathan, col. <<Fac-cinéma>>. Trad. Cast.: El tragaluz del infinito, Madrid, Cátedra, 1987. ISBN: 9788437606422
- [45] Bordwell, D. *Making Meaning: Inference and Rethoric in the Interpretation of Cinema*, Cambridge, Harvard University Press, 1989. Trad. Cast: El significado del filme. Inferencia y retórica en la interpretación cinematográfica, Barcelona, Paidós, 1995. ISBN: 067454336X
- [46] Font, D. *Paisajes de la modernidad. Cine europeo 1960-1980*. Ed. Paidós, Barcelona 2002. ISBN: 8449312256
- [47] Zunzunegui, S. *Pensar la imagen*. Cátedra. 2007. ISBN 9788437608150
- [48] www.cinemetrics.lv/ (Consultado el 20/09/2012)
- [49] www.dublincore.org/ (Consultado el 20/09/2012)
- [50] S. Manjunath, Philippe Salembier y Thomas Sikora; "Introduction to MPEG7: Multimedia Content Description interface" B. Ed. John Wiley & Sons, Ltd. 2002. ISBN: 0471486787
- [51] www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=34230 (Consultado el 20/09/2012)
- [52] www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=34231 (Consultado el 20/09/2012)
- [53] Eidenberg, H. "Distance measures for MPEG-7 based retrieval" *Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval*, Nov. 2003, pp. 130-137.
- [54] Brian Potetz, Tai Sing Lee, "Statistical correlation between two-dimensional images and three-dimensional structures in natural scenes". *Journal of the Optical Society of America A*, vol. 20, no 7, julio 2003.
- [55] Zhao Shan, Wang Hai-tao, "Image Retrieval Based on Bit-plane Distribution Entropy" *2008 International Conference on Computer Science and Software Engineering*. vol. 4, pp. 532-535, diciembre 2008
- [56] Zhitao Xiao, Jun Wu. "Analysis on image Symmetry Detection Algorithms" *IEEE Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007)*, vol.4, pp.745-750, Agosto 2007

- [57] Faustino Sánchez y Aarón Rodríguez. Retratos de familia – Tránsitos del cine. Infiel (Liv Ullman/Ingmar Bergman) y Yi Yi (Edward Yang). Editorial Shangrila, Santander, 2011. ISBN: 978-84-939366-0-0
- [58] Rosenbaum, Jonathan. “Elephant”. Filmkrant, 2004
www.jonathanrosenbaum.com/?p=20787 (Consultado el 20/09/2012)
- [59] Eisenstein, S. M, *Au-delà des étoiles*, Union générale d'éditions, 1974, París.
- [60] Bazin A., “Montage interdit”, *Qu'est-ce que le cinéma?* Ed. Du Cerf, París
Traducción castellana ¿Qué es el cine?, Ediciones Rialp, 1966, Madrid
- [61] J. Canny, “A computational approach to edge detection”. *IEEE PAMI* vol. 8, no. 6, págs. 679-698, Nov 86
- [62] Satoshi Suzuki, Keiichi Abe. “Topological structural analysis of digitized binary images by border following”. *Computer Vision, Graphics, and Image Processing* vol. 30, no 1, pp. 32-46, 1985
- [63] B. D. Lucas and T. Kanade (1981), “An iterative image registration technique with an application to stereo vision”. *Proceedings of Imaging Understanding Workshop*, pages 121—130
- [64] Ba Tu Truong; Venkatesh, S.; Dorai, C.; "Scene extraction in motion pictures," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.13, no.1, pp. 5- 15, Jan 2003
- [65] Tong Lin; Hong-Jiang Zhang; "Automatic video scene extraction by shot grouping," *Proceedings. 15th International Conference on Pattern Recognition*, vol.4, no., pp.39-42 vol.4, 2000
- [66] Jinhyung Cho; Kwiseok Kwon; Yongtae Park; , "Collaborative Filtering Using Dual Information Sources," *IEEE Intelligent Systems*, vol.22, no.3, pp.30-38, May-June 2007.
- [67] www.pintura.aut.org (Consultado el 20/09/2012)
- [68] www.grouplens.org/ (Consultado el 20/09/2012)
- [69] Ricardo Vélez Ibarrola, Alfonso García Pérez. *Principios de inferencia estadística. Cálculo de probabilidades y estadística matemática*. Universidad Nacional de Educación a Distancia, 1997. ISBN: 84-362-2947-9
- [70] Herlocker, J., Konstan, J. A., Borchers, A., Riedl, J.: “An Algorithm framework for performing collaborative filtering”. *Proceedings of SIGIR'99*, pp 230-237. 1999

- [71] Resnick, P., Iacovou, N. Suchak, M., Bergstrom, P., Riedl, J. "GroupLens: An open architecture for collaborative filtering of netnews". *Proceedings of ACM CSW'94 Conference on Computer Cooperative Work*. 1994
- [72] Shardanand, U., Maes, P.; "Social Information Filtering: Algorithms for automating 'word of mouth' ". *Proceedings of ACM CHI'95 Conference on Human Factors in Computing Systems*, pp. 210-217. 1995.
- [73] Balabanovic, M., Shoham, Y. 1997. "Fab: Content-based, collaborative recommendation". *Communications of ACM* vol. 40, no. 3, pp. 66-72, Marzo 1997
- [74] Ying Gao; Hong Qi; Jie Liu; Dayou Liu. "A Recommendation Algorithm Combining User Grade-Based Collaborative Filtering and Probabilistic Relational Models". *Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007)*, vol. 1, pp. 67-70
- [75] Breese, J. S., D. Heckerman, and C. Kadie. "Empirical analysis of predictive algorithms for collaborative filtering". *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, Madison, pp. 43-52, July 1998
- [76] Ponce Medellín Ismael Rafael. *Búsquedas Contextuales de Servicios Basados en Localización en un entorno de Web Social*, tesis de doctorado, Depto. de Ciencias Computacionales, Centro Nacional de Investigación y Desarrollo Tecnológico, Cuernavaca, Mor.
- [77] Laurent Candillier , Frank Meyer , Marc Boullé, "Comparing State-of-the-Art Collaborative Filtering Systems", *Proceedings of the 5th international conference on Machine Learning and Data Mining in Pattern Recognition*, pp. 548-562, July 18-20, 2007, Leipzig, Germany.
- [78] Aggarwal, C. C., J. L. Wolf, K-L. Wu, and P. S. Yu. "Horting hatches an egg: A new graphtheoretic approach to collaborative filtering". *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 201-212, August 1999.
- [79] Resnick, P., N. Iacovou, M. Sushak, P. Bergstrom, and J. Riedl. "GroupLens: An open architecture for collaborative filtering of netnews". *Proceedings of the 1994 Computer Supported Cooperative Work Conference*, pp. 175-186, 1994
- [80] Shardanand, U. and P. Maes. "Social information filtering: Algorithms for automating 'word of mouth' ". *Proc. of the Conf. on Human Factors in Computing Systems*, pp.210-217, 1995
- [81] Sarwar, B., G. Karypis, J. Konstan, and J. Riedl. "Item-based Collaborative Filtering Recommendation Algorithms". *Proc. of the 10th International WWW Conference*, pp. 285-295, 2001

- [82] M. Pazzani and D. Billsus, "Learning and Revising User Profiles: The Identification of Interesting Web Sites," *Machine Learning*, vol. 27, pp. 313-331, 1997.
- [83] R. Burke. "Knowledge-based recommender systems". *Encyclopedia of Library and Information Systems*, vol. 69, no.32, pp.175–186, 2000.
- [84] Boutemedjet, S.; Ziou, D.; "A Graphical Model for Context-Aware Visual Content Recommendation", *IEEE Transactions on Multimedia*, vol.10, no.1, pp.52-62, Jan. 2008
- [85] D. Maltz and K. Ehrlich. "Pointing the Way: Active Collaborative Filtering". In *Proceedings of the Conference on Human Factors in Computing Systems* (CHI95). ACM Press, 1995
- [86] R. Burke. "Integrating Knowledge-Based and Collaborative Filtering Recommender Systems". In *Proceedings of the Workshop on Artificial Intelligence for Electronic Commerce*, pp. 69–72. AAAI Press, 1999
- [87] Y. Freund, R. Iyer, R. Schapire, and Y. Singer. "An Efficient Boosting Algorithm for Combining Preferences". In *Proceedings of the Fifteenth International Conference on Machine Learning*, pp. 170–178. Morgan Kaufmann, 1998
- [88] C. Aggarwal, J. Wolf, K. Wu, and P. Yu. "Hortling Hatches an Egg: A New Graph-Theoretic Approach to Collaborative Filtering". In *KDD'99, Proceedings of the Fifth ACMS IKGDD International Conference on Knowledge Discovery and Data Mining*, pp. 201–212. ACM Press, 1999
- [89] B. Krulwich. "Lifestyle finder: Intelligent user profiling using large-scale demographic data". *AI magazine*, 18(2):37, 1997.
- [90] M.J. Pazzani. "A framework for collaborative, content-based and demographic filtering". *Artificial Intelligence Review*, 13(5):393–408, 1999
- [91] Littlestone, N. and Warmuth, M. "The Weighted Majority Algorithm". *Information and Computation* vol 108, no.2, pp. 212–261, 1994
- [92] Kazienko P., Musiał K., and Kajdanowicz T. "Multidimensional Social Network in the Social Recommender System". *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*. Vol. 41, N° 4, Julio 2011
- [93] Arazy O., Kumar N., Shapira B. "Improving Social Recommender Systems". *Social Computing. IEEE Computer Society*. Julio/agosto 2009
- [94] Granovetter, M.S. (1983). "The Strength of the Weak Tie: Revisited" [PDF], *Sociological Theory*, Vol. 1, 201-33.

- [95] Krackhardt, D. (1992) "The Strength of Strong Ties: The Importance of Philos in Organizations." In N. Nohria & R. Eccles (eds.), *Networks and Organizations: Structure, Form, and Action* pp. 216-239. Boston, MA: Harvard Business School Press.
- [96] Gilbert E., Karahalios K., "Predicting Tie Strength With Social Media". *Proceedings of the 27th international conference on Human factors in computing systems*. ACM, New York, 2009.
- [97] Pazzani, M.J., Billsus, D.: "Content-based recommendation systems". Brusilovsky, P., Kobsa, A., Neidl, W. (eds.): *The Adaptive Web: Methods and Strategies of Web Personalization. Lecture Notes in Computer Science*, Vol. 4321. Springer-Verlag, Berlin Heidelberg New York (2007).
- [98] Peng, Tao; Wang, Wendong; Gong, XiangYang; Tian, Ye; Yang, XiaoGang; Ma, Jian. "A Graph Indexing Approach for Content-Based Recommendation System". *2010 Second International Conference on Multimedia and Information Technology* (MMIT), vol.1, pp. 93-97, April 2010
- [99] J.J. Rocchio, "Relevance Feedback in Information Retrieval" SMART Retrieval System—*Experiments in Automatic Document Processing*, G. Salton, ed., chapter 14, Prentice Hall, 1971.
- [100] Domingos, Pedro and Michael Pazzani (1997) "On the optimality of the simple Bayesian classifier under zero-one loss". *Machine Learning*, 29:103–137
- [101] R. Baeza-Yates and B. Ribeiro-Neto, "Modern Information Retrieval". Addison-Wesley, 1999.
- [102] M. Pazzani and D. Billsus, "Learning and Revising User Profiles: The Identification of Interesting Web Sites," *Machine Learning*, vol. 27, pp. 313-331, 1997.
- [103] Sebastián Pizard. ClaNFi, "Clasificador de Noticias Financieras". Instituto de Computación. Facultad de Ingeniería. Universidad de la República Oriental del Uruguay. Agosto 2007. www.fing.edu.uy/inco/grupos/pln/prygrado/InformeCLANFI.pdf (Consultado el 20/09/2012)
- [104] Sangoh Jeong; Kalasapur, S.; Cheng, D.; Song, H.; Hyuk Cho; "Clustering and Naïve Bayesian Approaches for Situation-Aware Recommendation on Mobile Devices," *International Conference on Machine Learning and Applications*, 2009. ICMLA '09., pp.353-358, 13-15 Dec. 2009
- [105] Dong-Moon Kim; Kun-su Kim; Kyo-Hyun Park; Jee-Hyong Lee; Keon Myung Lee; "A music recommendation system with a dynamic k-means clustering

algorithm,” *Machine Learning and Applications*, 2007. ICMLA 2007. Sixth International Conference on, pp.399-403, 13-15 Dec. 2007

- [106] Teuvo Kohonen, Samuel Kaski, Krista Lagus, Jarkko Salojärvi, Vesa Paatero, and Antti Saarela. “Organization of a massive document collection”. *IEEE Transactions on Neural Networks*, Special Issue on Neural Networks for Data Mining and Knowledge Discovery, 11(3):574–585, May 2000.
- [107] www.websom.hut.fi/websom/ (Consultado el 20/09/2012)
- [108] Bhatt, R.B.; , “Neuro-fuzzy decision trees for content popularity model and multi-genre movie recommendation system over social network,” *TENCON 2009 - 2009 IEEE Region 10 Conference*, pp.1-6, 23-26 Jan. 2009
- [109] Li, P.; Yamada, S.; “A movie recommender system based on inductive learning,” *2004 IEEE Conference on Cybernetics and Intelligent Systems*, vol.1, pp. 318-323 vol.1, 1-3 Dec. 2004
- [110] www.showroom.daedalus.es/es/buscarte/infobuscarte.php (Consultado el 20/09/2012)
- [111] César Pérez López, Daniel Santín González; “Minería de datos. Técnicas y herramientas”; International Thomson Ediciones Paraninfo, S.A., 2ª edición, 2008, ISBN: 978-84-9732-492-2.
- [112] R. C. Veltkamp and M. Tanase. “Content-Based Image Retrieval Systems: A Survey”. Technical Report UU-CS-2000-34, Dept. of Computing Science, Utrecht University, 2000.
- [113] Chan Young Kim; Jae Kyu Lee; Yoon Ho Cho; Deok Hwan Kim; “VISCORS: a visual-content recommender for the mobile Web”, *IEEE Intelligent Systems*, vol.19, no.6, pp. 32- 39, Nov.-Dec. 2004
- [114] Yong Rui; Huang, T.S.; Ortega, M.; Mehrotra, S.; “Relevance feedback: a power tool for interactive content-based image retrieval”, *IEEE Transactions on Circuits and Systems for Video Technology*, vol.8, no.5, pp.644-655, Sep 1998
- [115] Tatemura, J.; Santini, S.; Jain, R.; “Social and content-based information filtering for a Web graphics recommender system”, *Proceedings. International Conference on Image Analysis and Processing*, pp.842-847, 1999
- [116] Boutemedjet, S.; Ziou, D.; “A Graphical Model for Context-Aware Visual Content Recommendation”, *IEEE Transactions on Multimedia*, vol.10, no.1, pp.52-62, Jan. 2008
- [117] Hanjalic, A.; Li-Qun Xu; “Affective video content representation and modeling”, *IEEE Transactions on Multimedia*, vol.7, no.1, pp. 143- 154, Feb. 2005

- [118] Calcanis, C.; Callaghan, V.; Gardner, M.; Walker, M.; “Towards end-user physiological profiling for video recommendation engines”, *2008 IET 4th International Conference on Intelligent Environments*, pp.1-5, 21-22 July 2008
- [119] Chi-Cheng Tsai; Ching-I Chung; Yi-Ting Huang; Chia-Hsing Shen; Yu-Chieh Wu; Jie-Chi Yang; “VCSR: Video Content Summarization for Recommendation”, *Seventh IEEE International Conference on Advanced Learning Technologies*, 2007. ICAALT 2007., pp.862-864, 18-20 July 2007
- [120] Xiao Wu; Yongdong Zhang; Junbo Guo; Jintao Li; “Web video recommendation and long tail discovering”, *2008 IEEE International Conference on Multimedia and Expo*, pp.369-372, June 23 2008- June 26 2008
- [121] Zhang, Q.; Izquierdo, E.; “Bayesian learning and reasoning for context exploitation in visual information retrieval”, *5th International Conference on Visual Information Engineering*, 2008. VIE 2008., vol., no., pp.170-175, July 29 2008-Aug. 1 2008
- [122] Feng Gao; Yuhong Li; Li Han; Jian Ma; InfoSlim: “An Ontology-Content Based Personalized Mobile News Recommendation System”, *5th International Conference on Wireless Communications, Networking and Mobile Computing*, 2009. WiCom '09., pp.1-4, 24-26 Sept. 2009
- [123] Lee, J.S.A.; Jayant, N.; “Mixed-initiative multimedia for mobile devices: Design of a semantically-relevant low-latency system for news video recommendations”, *Southeastcon, 2009*. IEEE, pp.267-272, 5-8 March 2009
- [124] Bianco-Fernandez, Y.; Pazos-Arias, J.J.; Gil-Solla, A.; Ramos-Cabrera, M.; Lopez-Nores, M.; “Providing Entertainment by Content-based Filtering and Reasoning in Intelligent Recommender Systems”, *Digest of Technical Papers. International Conference on Consumer Electronics*, 2008. ICCE 2008., pp.1-2, 9-13 Jan. 2008
- [125] Park, J; Lee, S; Kim, K; Chung, B; Lee, Y; “An Online Video Recommendation Framework Using View Based Tag Cloud Aggregation”, *IEEE Multimedia*, no.99, pp.1-1.
- [126] Nai-Xiang Lian; Yap-Peng Tan; “Probabilistic approach to K-nearest neighbor video retrieval”, *Proceedings of the 2004 International Symposium on Circuits and Systems*, 2004. ISCAS '04., vol.2, no., pp. II- 193-6 Vol.2, 23-26 May 2004
- [127] Claypool, M., A. Gokhale, T. Miranda, P. Murnikov, D. Netes, and M. Sartin. “Combining content-based and collaborative filters in an online newspaper”. In *ACM SIGIR'99. Workshop on Recommender Systems: Algorithms and Evaluation*, August 1999

- [128] Pazzani, M. "A framework for collaborative, content-based and demographic filtering". *Artificial Intelligence Review*, pages 393-408, December 1999
- [129] Billsus, D. and M. Pazzani. "User modeling for adaptive news access". *User Modeling and User-Adapted Interaction*, vol.10, no.2-3, pp.147-180, 2000
- [130] Melville, P., R. J. Mooney, and R. Nagarajan. "Content-Boosted Collaborative Filtering for Improved Recommendations". In *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, pp. 187-192, Edmonton, Canada, 2002
- [131] Soboroff, I. and C. Nicholas. "Combining content and collaboration in text filtering". In *IJCAI'99 Workshop: Machine Learning for Information Filtering*, pp.86-91, August 1999
- [132] W.W. Cohen, R.E. Schapire, and Y. Singer. "Learning to order things". *Journal of Artificial Intelligence Research*, vol.10, pp.243–270, 1999.
- [133] P. Cotter and B. Smyth. PTV, "Intelligent personalised TV guides". In *proceedings of the national conference on artificial intelligence*, pp. 957–964. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2000.
- [134] A. Wasfi and M. Ahmad. "Collecting user access patterns for building user profiles and collaborative filtering". In *Proceedings of the 4th international conference on Intelligent user interfaces*, pp. 57–64. ACM, 1998.
- [135] R. Burke. "Hybrid recommender systems: Survey and experiments". *User Modeling and User-Adapted Interaction*, vol.12, no.4, pp.331–370, 2002
- [136] Basu, C., H. Hirsh, and W. Cohen. "Recommendation as classification: Using social and content-based information in recommendation". In *Recommender Systems. Papers from 1998 Workshop*. Technical Report WS-98-08. AAAI Press, 1998
- [137] Tommaso Di Noia, Eugenio Di Sciascio, and Francesco M. Donini. "Semantic matchmaking as non-monotonic reasoning: A description logic approach". *Journal of Artificial Intelligence Research*, vol.29, pp.269–307, 2007
- [138] N. Littlestone and M.K. Warmuth. The weighted majority algorithm, *Information and Computation Journal*, vol.108, no.2, pp.212-261, 1989.
- [139] Popescul, A., L. H. Ungar, D. M. Pennock, and S. Lawrence. "Probabilistic Models for Unified Collaborative and Content-Based Recommendation in Sparse-Data Environments". In *Proc. of the 17th Conf. on Uncertainty in Artificial Intelligence*, Seattle, WA, pp. 437-444, 2001

- [140] Schein, A. I., A. Popescul, L. H. Ungar, and D. M. Pennock. "Methods and metrics for cold-start recommendations". *In Proc. of the 25th Annual Intl. ACM SIGIR Conf.*, pp.253-260, 2002
- [141] Hofmann, T. "Probabilistic Latent Semantic Analysis". *In Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pp. 289-296, 1999
- [142] Condliff, M., D. Lewis, D. Madigan, and C. Posse. "Bayesian mixed-effects models for recommender systems". *In ACM SIGIR'99 Workshop on Recommender Systems: Algorithms and Evaluation*, August 1999
- [143] Ansari, A., S. Essegai, and R. Kohli. "Internet recommendations systems". *Journal of Marketing Research*, pp. 363-375, August 2000
- [144] J. L. Herlocker, J. A. Constant, L. G. Terveen, J. T. Riedl. "Evaluating Collaborative Filtering Recommender Systems". *ACM Transactions on Information Systems*, Vol. 22, N° 1, January 2004, Pages 5-53.
- [145] J. O'Donovan, B. Smyth. "Trust in Recommender Systems". *IUI'05, Proceedings of the 10th international conference on Intelligent user interfaces*. January 9–12, 2005, pp.167-174, San Diego, California, USA.
- [146] Tintarev, N.; Masthoff, J.; , "A Survey of Explanations in Recommender Systems," *IEEE 23rd International Conference on Data Engineering Workshop*, 2007, pp.801-810, 17-20 April 2007
- [147] M. G. Vozalis and K. G. Margaritis, "A recommender system using Principal Component Analysis", *Current Trends in Informatics*, Vol. 1, New Technologies Publications, pp. 271–283, 2007
- [148] Trevor F. Cox, Michael A. A. Cox, "Multidimensional Scaling". *Monographs on Statistics and Applied Probability* 88. 2nd ed. ISBN: 1-58488-094-5
- [149] Hadjahmadi, A.H.; Homayounpour, M.M.; Ahadi, S.M.; "Robust weighted fuzzy c-means clustering," *IEEE International Conference on Fuzzy Systems*, 2008. FUZZ-IEEE 2008. (IEEE World Congress on Computational Intelligence), pp.305-311, 1-6 June 2008
- [150] Sledge, I.J.; Bezdek, J.C.; Havens, T.C.; Keller, J.M.; , "Relational Generalizations of Cluster Validity Indices," *IEEE Transactions on Fuzzy Systems*, vol.18, no.4, pp.771-786, Aug. 2010
- [151] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 8, pp. 841–847, Aug. 1991

- [152] Jones, B.C.; Wilkes, D.M.; "A new analysis framework for relevance feedback-driven similarity measure refinement in content-based image retrieval," *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001. CVPR 2001., vol.1, no., pp. I-920- I-925 vol.1, 2001
- [153] Sugimoto, O.; Naito, S.; Sakazawa, S.; Koike, A.; "Objective perceptual video quality measurement method based on hybrid no reference framework," 16th IEEE International Conference on Image Processing (ICIP), 2009, pp.2237-2240, 7-10 Nov. 2009
- [154] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," in *Proc. 14th Conf. Uncertainty Artificial Intelligence*, pp. 43–52, 1998
- [155] R. M. Bell, Y. Koren, and C. Volinsky. *The BellKor 2008 solution to the Netflix Prize*. Disponible en www.netflixprize.com, 2008. (Consultado el 20/09/2012)
- [156] Jähne B.; "Practical handbook on Image Processing for Scientific and Technical Applications". CRC Press, 2004, 2ª Edición. ISBN: 0-8493-1900-5
- [157] Suzuki S., Abe K.; "Topological Structural Analysis of Digitized Binary Images by Border Following"; *Computer Vision, Graphics, and Image Processing* 30, pp. 32-46; 1985.
- [158] Lucas B., Kanade T.; "An iterative image registration technique with an application to stereo vision". *Proceedings of Imaging Understanding Workshop*, pages 121—130; 1981.

Acrónimos

3SME	3.0 Social Media Ecosystem
ACL	Average Scene Length
ALI	Attribute Local Interest
AR	Attribute Relevance
ASL	Average Shot Length
AVC	Attribute Value Certainty
BWR	Black and White Rate
CBIR	Content-Based Image Retrieval
CE	Coherencia Espacial
CIV	Camera motion Intensity Variance
CLD	Color Layout Descriptor
CMI	Camera Motion Intensity mean
CMC	Camera Motion Complexity mean
CMV	Camera Motion Complexity variance
CNN	Cable News Network
CRV	Chromatic Variety
CV	Coefficient of Variation
DCD	Dominant Color Descriptor
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DIR	Dissolve Rate
EDL	Energía de líneas
EHD	Edge Histogram Descriptor
ENM	Entropía Media
EPB	Entropía por Planos de Bits
FAR	Fade Rate
FFT	Fast Fourier Transform
HbbTV	Hybrid Broadcast Broadband TV
HDL	Homogeneidad De Líneas
HMM	Hidden Markov Model
HSL	Hue-Saturation-Lightness
ICV	Internal Motion Complexity Variance
IEC	International Electrotechnical Commission
IIV	Internal Intensity Variance
IMC	Internal Motion Complexity mean
IMI	Internal Motion Intensity mean
ISO	International Organization for Standardization
JOI	Juegos Olímpicos Interactivos
LEN	Length of the film
LM	Luminancia Media
LSI	Latent Semantic Indexing
LUM	Luminosidad Media

LUV	Luminance Variance
MAE	Mean Absolute Error
MDS	Multidimensional Scaling
MHP	Multimedia Home Platform
MIREIA	Distribución de contenidos mediante Motores Inteligentes de Recomendación Adaptados a Redes Sociales
MPEG	Moving Picture Experts Group
MSL	Median Shot Length
MySQL	My Structured Query Language
NoS	Number of Shots
NR	Número de Recomendaciones
PC	Personal Computer
PC	Planaridad de color
PCA	Principal Component Analysis
PDA	Personal Digital Assistant
PTV	Personal Television
PVR	Personal Video Recorder
QoE	Quality of Experience
RGB	Red-Green-Blue
RMSE	Root Mean Square Error
RR	Recommendation Relevance
SA	Saturación
SAT	Saturación
SCR	Shot/Scene Rate
SOM	Self Organizing Maps
SPR	Speech Rate
SIR	Silence Rate
TF-IDF	Term Frequency/Inverse Document Frequency
TV	Television
TVE	Televisión Española
UA	Usuarios Afectados
VC	Variedad Cromática
VCL	Variance Scene Length
VE	Variedad de Entropía
VSC	Variance of Shot/Scene rate
VSL	Variety of Shot Length
VTL	Video Total Length
WEBSOM	Web Self-Organizing Map
XML	eXtensible Markup Language
XOR	eXclusive OR

ANEXO I: MODELO DE ENCUESTA PARA LAS PRUEBAS DE MODELADO DE CONSUMO

Nombre:

Ordene los siguientes temas por interés (1 el mayor interés, 10 el menos) y puntúelos del 1 al 5 (1 no me gusta nada, 2 no me gusta, 3 me da lo mismo, 4 me gusta, 5 me gusta mucho):

1. Corazón
2. Series
3. Informativos
4. Deportes
5. Cine
6. Documentales
7. Concursos
8. Reality Shows
9. Infantil
10. Magazine

Puntúe del 1 al 5 los siguientes programas, y con un 0 si no los conoce o no los ha visto: (1 no me gusta nada, 2 no me gusta, 3 me da lo mismo, 4 me gusta, 5 me gusta mucho)

PROGRAMAS

- | | |
|---------------------------------|---------------------------|
| 1. Agrosfera | 20. Informe semanal |
| 2. Ahora caigo | 21. La casa de al lado |
| 3. Al rojo vivo | 22. La casa de empeños |
| 4. Amar en tiempos revueltos | 23. La fuerza del destino |
| 5. Balonmano | 24. La hora de Jose Mota |
| 6. Cazadores de nazis | 25. La mañana de La 1 |
| 7. Cielo rojo | 26. La sala |
| 8. Cine de barrio | 27. Locos por ganar |
| 9. Corazón (La 1) | 28. Megatrix |
| 10. Dale al Rec | 29. Noticias (La Sexta) |
| 11. Deportes Cuatro | 30. Noticias (Cuatro) |
| 12. El documental (La 2) | 31. Otra movida |
| 13. El secreto de Puente Viejo | 32. Pura sangre |
| 14. Futbol | 33. Sálvame |
| 15. Gente | 34. Telediario (La 1) |
| 16. Gran hermano | |
| 17. Grandes documentales (La 2) | |
| 18. Informativos Telecinco | |
| 19. Noticias (Antena 3) | |

ANEXO II: ALGORITMOS EMPLEADOS PARA LA OBTENCIÓN DE LOS DESCRIPTORES DE BAJO NIVEL

Binarización de una imagen mediante el filtro de Canny

El filtro de detección de bordes de Canny es un proceso multifase, en el que el usuario introduce tres parámetros básicos: una desviación estándar y dos valores umbral. El resultado es una imagen binaria, en la que aparecerán en negro los píxeles correspondientes a los bordes y en blanco el resto de píxeles, o viceversa.

A continuación se describen cada una de las fases que intervienen en la aplicación del filtro de Canny, que se basa en las derivadas de la gaussiana [156]:

Fase 1. Suavizado de la imagen. En primer lugar, la imagen original se suaviza mediante la convolución con una función gaussiana (124) de anchura especificada por el usuario, con el objetivo de atenuar posibles ruidos de la imagen. La desviación estándar considerada como parámetro de entrada en el filtro de Canny define el tamaño o anchura de la máscara gaussiana.

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (124)$$

Fase 2. Realce de la imagen. La imagen suavizada en la primera fase es convolucionada mediante un operador gradiente en las direcciones de los ejes X e Y, de forma que se obtenga una imagen con los cambios de intensidad detectados sobre la que calcular los componentes del borde relativos a la fuerza (125) y orientación (126) de la normal al borde. La fuerza del borde proporciona una medida del contraste local de la imagen, mientras que la orientación del borde identifica la dirección perpendicular al borde en forma de vector unitario.

$$e_s(i, j) = \sqrt{J_x^2(i, j) + J_y^2(i, j)} \quad (125)$$

$$e_o(i, j) = \arctan \frac{J_y}{J_x} \quad (126)$$

Fase 3. Eliminación de los no máximos. Sobre la imagen realizada con los cambios de intensidad de la fase anterior, se encuentra la dirección que mejor se aproxima a la dirección normal al borde. Si la fuerza del borde es menor que al menos uno de los dos vecinos a lo largo de esa dirección se suprime el punto como punto perteneciente a un borde. El resultado es una imagen con los bordes adelgazados como consecuencia de la supresión.

Fase 4. Umbralización de bordes. La asignación del umbral es quizás la fase más delicada del filtro de Canny, ya que se debe evitar la posible presencia de bordes ruidosos. La mayoría de los detectores de bordes emplean un simple umbral, ocasionando la partición de las líneas siempre que los valores fluctúen arriba y abajo del valor umbral.

Algoritmo de Suzuki [157]. Extracción del número de bordes

El algoritmo de Suzuki se utiliza, una vez que la imagen ha sido binarizada mediante el filtro de Canny, para determinar su número de bordes. Las imágenes binarias se caracterizan por tener sólo dos tipos de píxeles, de valor 0 (0-píxeles) o 1 (1-píxeles). Un conjunto de 0-píxeles o 1-píxeles conectados forman una componente. Para ello, se definen dos tipos de conectividad: la 4-conectividad y la 8-conectividad.

Dos píxeles con coordenadas (x', y') y (x'', y'') son 4-conexos si y sólo si:

$$|x' - x''| + |y' - y''| = 1 \quad (127)$$

Y son 8-conexos si y sólo si:

$$\max(|x' - x''|, |y' - y''|) = 1 \quad (128)$$

Utilizando este concepto de conectividad, se puede segmentar una imagen binaria en componentes 4-conexas y 8-conexas formadas por 0-píxeles y 1-píxeles. Cada una de estas componentes se compone de píxeles de valores iguales y, para cualquier par de píxeles, existe un 4-camino u 8-camino¹⁴ que los conecta.

Ya que las 1-componentes (componentes que contienen 1-píxeles) son complementarias con las 0-componentes, se puede considerar únicamente las primeras y asumir que todo lo demás corresponde a 0-píxeles que conforman el fondo de la imagen. Cada 1-componente posee un único borde exterior que lo separa de la 0-componente que lo rodea y 0 o más bordes interiores que lo separan de las 0-componentes que rodea (agujeros).

El objetivo de este algoritmo es entonces recolectar todos los puntos de la imagen caracterizados como puntos de borde. Para hacer esto, recorre la imagen línea por línea buscando nuevos puntos de borde; cuando encuentra uno nuevo, activa un proceso de seguimiento de borde almacenando las coordenadas de cada punto que lo compone. Durante este proceso va marcando los píxeles ya visitados asignándoles a estos un valor especial para no volver a incidir en ellos. Finalmente se disponen todos los contornos encontrados en una lista y se realiza el recuento final para el descriptor o las operaciones necesarias en los algoritmos.

Algoritmo de Lucas-Kanade [158] para la detección del flujo de movimiento

El método de Lucas-Kanade asume que el desplazamiento de la imagen entre dos fotogramas cercanos es pequeño y prácticamente constante en un entorno de cada punto que se considere. Si llamamos Ω_p al conjunto de coordenadas que forman un entorno del píxel p , para cada $q \in \Omega_p$ debe verificarse que:

$$\nabla I^T \cdot V = \frac{\partial I}{\partial t}(q) \quad (129)$$

¹⁴ Un 4-camino u 8-camino, es una secuencia de píxeles tal que cada píxel es 4-conexo u 8-conexo con el píxel siguiente.

Expresando la ecuación del problema de la apertura en cada uno de los $n = |\Omega_p|$ píxeles que forman un entorno de p , se obtiene la ecuación lineal $Av=b$, donde:

$$A = \begin{pmatrix} I_x(q_1) & I_y(q_1) \\ \vdots & \vdots \\ I_x(q_n) & I_y(q_n) \end{pmatrix}, v = \begin{pmatrix} u \\ v \end{pmatrix}, y b = \begin{pmatrix} -I_t(q_1) \\ \vdots \\ -I_t(q_n) \end{pmatrix} \quad (130)$$

Siendo $I_x = \frac{\partial I}{\partial x}$, $I_y = \frac{\partial I}{\partial y}$ e $I_t = \frac{\partial I}{\partial t}$

Dado que la ecuación tiene más ecuaciones que incógnitas, el método de Lucas-Kanade consiste en obtener una solución aplicando el método de los mínimos cuadrados, lo que consigue resolviendo la ecuación $A^T A v = A^T b$ obteniendo como solución $v = (A^T A)^{-1} A^T b$.

El método otorga igual importancia a todos los píxeles que forman el entorno de p . Si W es una matriz diagonal $n \times n$, donde el elemento (i,i) es el peso de $q_i \in \Omega_p$, entonces la ecuación $A^T W A v = A^T W b$ resuelve el problema de la apertura dando, ahora sí, distinto nivel de importancia a los distintos píxeles que forman el entorno del punto estudiado. En la práctica, se suelen usar pesos que varían según una distribución normal en función de la distancia del punto q_i a p .

ANEXO III: CORPUS SELECCIONADO PARA EL RECOMENDADOR DE IMÁGENES

Id	Nombre
1	Lavabo
2	Lying Figure
3	Katia Lisant
4	Los esposos
5	Naturaleza muerta con Girasol
6	Helios and Phaeton with Saturn and Four Seasons
7	Mujer en el jardín de Saint-Adresse
8	Allegorical Portrait of Philip IV
9	El inconsciente azul
10	La batalla de Trafalgar vista desde las jarcias del palo de mesana del Vistoy
11	La hilandera, la cabrera de Auvernia
12	Cristo de San Juan de la Cruz
13	Girl in striped nightdress
14	Portrait of an Artist
15	Venice: Entrance to the Grand Canal
16	Portrait of M. Pertuiset, the Lion Hunter
17	Sinagoga de Hadassa. "La tribu de Daniel"
18	Flora
19	Circuncisión
20	El canto del ruiseñor a media noche y la lluvia matinal
21	The Bathers
22	Retrato de Madame de Senonnes
23	Mujer vuelta hacia la derecha
24	Arearea
25	Lady Standing at the Virginals
26	L'incubo abbandona il giaciglio di due fanciulle dormienti
27	Papel de empapelar con vacas
28	Adán y Eva
29	Matin à venise
30	Violeta
31	Farm Garden with Sunflowers
32	San Francisco de Borja se despide de su familia
33	L'Odalisque à la culotte rouge
34	Self-Portrait with Soft Hat
35	Junto al lecho de muerte
36	Interior (with piano and woman in black, Strandgade 30)
37	La bacanal de los Andrios
38	Ciudad amarilla
39	François Duquesnoy
40	Drugstore
41	Valencia, dos niños en una playa

42	Camino con sauces podados
43	San Juan Bautista
44	Still Life with Goldfish Bowl
45	Sin título
46	Birds´ Hell
47	Lumière de soir, près de Vernon
48	La incertidumbre del poeta
49	El Jardín de las Delicias. Panel Central
50	Ground swell
51	La Gran Vía
52	Madrid desde Capitán Haya
53	Nu devant la cheminée
54	Anna Zborowska
55	Composition with Red. Yellow and Blue
56	The dance to the music of time
57	Camille Monet and Her Son Jean
58	Vieja friendo huevos
59	Convergence
60	El barco negrero
61	Las Espigadoras
62	La Tentación de San Antonio
63	Reflexión (Autorretrato)
64	The painter's mother resting, II
65	La entrada del arsenal
66	La amazona de frente
67	Magdalena penitente
68	Baltasar ve la inscripción en la pared
69	El bello pájaro descifrando lo desconocido a una pareja de enamorados
70	The Seine at Le Grande Jatte
71	Portrait of Delphine Ingres, née Ramel
72	Retrato de Dora Maar
73	Tahitian Women with Mango Blossoms
74	El ama y su criada
75	El despertar de Titania. Rodeada de su cortejo, la reina de las hadas, siempre..
76	Mao Tse Tung
77	Susana en el baño
78	Le Ban de Diane
79	In the Blue
80	Water Serpents II
81	La lucha con los mamelucos
82	La danza
83	El fumador
84	Melancolía
85	White Doors or Open Doors (Strandgade 30)
86	El amor de una Romana por su padre
87	Embrace Lovers II
88	Autorretrato
89	Holland Hotel

90	El baño
91	La iglesia de Auvers
92	Partida de cartas
93	In the Car
94	All one-, two-, three-, and four-part combinations of gray. Yellow, red and blue
95	The Prodigal Son
96	La Palma
97	El gran juego
98	El Carro de Heno
99	Nighthawks
100	Study of Red Pope
101	La partida de naipes
102	Desnudo acostado con el brazo derecho bajo la cabeza
103	Cuadro II
104	El Rapto de las Sabinas
105	Waterlilies, Green Reflection
106	La túnica de José
107	Alchemy
108	Régulo
109	Planting Potatoes
110	Juego lúgubre
111	Mr. and Mrs. Clark and Percy
112	Regata en el Gran Canal
113	En el invernadero
114	Madonna of the Village
115	Danae
116	Betsabé con la carta de David
117	Constellation: Awakening at Dawn
118	The Models
119	Portrait of Vicomtesse Louise-Alvertine d'Haussonville
120	Retrato de Marie-Thérèse Walter
121	Dulces fantasías
122	Mujer tocando la guitarra
123	Pesadilla Nocturna
124	Liz
125	Leda y el cisne
126	Una mañana
127	Transverse Line
128	Tree of Life, Central Section , Pattern for Stoclet Frieze
129	Las lavanderas
130	Armonía en rojo
131	Bodegón con pañería
132	Vampire
133	Interior with Punch Bowl (Strandgade 30)
134	El jardín del amor
135	Desnudo femenino agazapado, con la mejilla apoyada en la...
136	Diana Cecil, condesa de Oxford
137	Canadian Club

138	El baño del caballo
139	Campo con campesino arando y molino
140	Decapitación de Holofernes
141	Mujer en el baño
142	Irregular horizontal bands of color superimposed
143	La noche
144	La Salle à manger à la campagne
145	La recompensa del Adivino
146	El Juicio Final. Panel Central
147	Morning Sun
148	Gran Via
149	Jet of Water
150	The Living Room
151	Paisaje
152	Composition II With Black lines
153	Paisaje con edificios
154	The Red Kerchief: Portrait of Camille Monet
155	Las Hilanderas
156	Composition with Pouring II
157	Venecia desde el pórtico de la iglesia de Salute
158	La lavandera
159	Dalí de espaldas pintando a Gala de espaldas...
160	Retrato del Barón H. H. Thyssen-Bornemisza
161	Olympic Blvd
162	Recepción del embajador imperial
163	Patinando
164	Memory of Paris
165	Alegoría de las Tres Edades de la Vida
166	Christ and the Woman of Samaria
167	Danger
168	The Channel at Gravelines, in the direction of the Sea
169	Pauline Eleanore de Galard de Brassac de Bearn
170	Woman with Book (Portrait of Marie-Thérèse Walter)
171	Tahitianas (En la playa)
172	Mujer sentada ante clavicémbalo
173	Shakespeare. Henry V, Act II, Scene II
174	Flores
175	Judith y Holofernes
176	Le Pont de Mantes
177	Sin título
178	Apple Tree I
179	Muchachos cogiendo fruta
180	Interior with Eggplants
181	Naturaleza muerta con Manzanas y Naranjas
182	The Dead Mother
183	Interior (with young woman seen from behind, Strandgade 30)
184	El protector de las Artes
185	La media verde

186	Queen Henrietta Maria
187	Cafe Express
188	Playa de San Sebastian
189	Campo de Trigo con Cipreses
190	San Jerónimo penitente
191	Sandwich and Soda from the portfolio Ten Works by Ten Painters
192	Irregular form with color bands
193	Large Still Life, Blue Interior
194	Le Pont des arts
195	Entrevista
196	Tríptico de la Adoración de los Magos
197	Verano
198	Sin título
199	Clavel
200	El baile de las Ninfas

ANEXO IV: CORPUS SELECCIONADO PARA EL RECOMENDADOR DE VÍDEOS

Id	Película	Número de valoraciones
16	Casino (1995)	9621
47	Seven (a.k.a. Se7en) (1995)	22521
50	Usual Suspects, The (1995)	24037
105	Bridges of Madison County, The (1995)	5676
111	Taxi Driver (1976)	12865
138	Neon Bible, The (1995)	93
190	Safe (1995)	610
194	Smoke (1995)	3664
215	Before Sunrise (1995)	2665
235	Ed Wood (1994)	9738
348	Bullets Over Broadway (1994)	5076
357	Four Weddings and a Funeral (1994)	16523
367	Mask, The (1994)	19036
412	Age of Innocence, The (1993)	2781
441	Dazed and Confused (1993)	4162
459	Getaway, The (1994)	697
480	Jurassic Park (1993)	32631
539	Sleepless in Seattle (1993)	16925
541	Blade Runner (1982)	15471
593	Silence of the Lambs, The (1991)	33668
597	Pretty Woman (1990)	19229
608	Fargo (1996)	23794
645	Nelly & Monsieur Arnaud (1995)	209
702	Faces (1968)	209
733	Rock, The (1996)	17964
778	Trainspotting (1996)	11791
841	Eyes Without a Face (Les Yeux sans visage) (1959)	209
858	Godfather, The (1972)	19814
894	Liebelei (1933)	14
903	Vertigo (1958)	7491
904	Rear Window (1954)	8825
911	Charade (1963)	2137
912	Casablanca (1942)	12507
923	Citizen Kane (1941)	9605
924	2001: A Space Odyssey (1968)	13061
928	Rebecca (1940)	2547
936	Ninotchka (1939)	813
942	Laura (1944)	1478
953	It's a Wonderful Life (1946)	7485
1084	Bonnie and Clyde (1967)	5651
1086	Dial M for Murder (1954)	2988

1089	Reservoir Dogs (1992)	13428
1090	Platoon (1986)	8721
1193	One Flew Over the Cuckoo's Nest (1975)	14435
1203	12 Angry Men (1957)	5365
1213	Goodfellas (1990)	13188
1219	Psycho (1960)	10384
1230	Annie Hall (1977)	8405
1232	Stalker (1979)	587
1234	Sting, The (1973)	8179
1244	Manhattan (1979)	4238
1246	Dead Poets Society (1989)	10604
1258	Shining, The (1980)	11423
1262	Great Escape, The (1963)	5597
1265	Groundhog Day (1993)	16743
1266	Unforgiven (1992)	7799
1333	Birds, The (1963)	5955
1339	Dracula (Bram Stoker's Dracula) (1992)	5274
1343	Cape Fear (1991)	5147
1348	Nosferatu (Nosferatu, eine Symphonie des Grauens) (1922)	1900
1464	Lost Highway (1997)	2087
1488	Devil's Own, The (1997)	2148
1625	Game, The (1997)	7898
1645	Devil's Advocate, The (1997)	6169
1682	Truman Show, The (1998)	11597
2126	Snake Eyes (1998)	2691
2160	Rosemary's Baby (1968)	3795
2167	Blade (1998)	6140
2174	Beetlejuice (1988)	10849
2176	Rope (1948)	1389
2231	Rounders (1998)	3419
2291	Edward Scissorhands (1990)	10623
2396	Shakespeare in Love (1998)	14121
2403	First Blood (Rambo: First Blood) (1982)	2689
2427	Thin Red Line, The (1998)	4348
2571	Matrix, The (1999)	23229
2628	Star Wars: Episode I - The Phantom Menace (1999)	15744
2671	Notting Hill (1999)	6762
2705	Late August, Early September (Fin août, début septembre) (1998)	63
2706	American Pie (1999)	3869
2712	Eyes Wide Shut (1999)	8138
2730	Barry Lyndon (1975)	1140
2731	400 Blows, The (Les Quatre cents coups) (1959)	1330
2762	Sixth Sense, The (1999)	19400
2763	Thomas Crown Affair, The (1999)	6690
2858	American Beauty (1999)	22120
2927	Brief Encounter (1946)	432

2947	Goldfinger (1964)	5946
2973	Crimes and Misdemeanors (1989)	2237
2997	Being John Malkovich (1999)	13871
3081	Sleepy Hollow (1999)	7139
3083	All About My Mother (Todo sobre mi madre) (1999)	2215
3224	Woman in the Dunes (Suna no onna) (1964)	315
3232	Seven Chances (1925)	63
3307	City Lights (1931)	1210
3408	Erin Brockovich (2000)	8786
3415	Mirror, The (Zerkalo) (1975)	185
3424	Do the Right Thing (1989)	3145
3435	Double Indemnity (1944)	2403
3451	Guess Who's Coming to Dinner (1967)	1444
3462	Modern Times (1936)	1663
3629	Gold Rush, The (1925)	1025
3641	Woman of Paris, A (1923)	32
3645	Cleo From 5 to 7 (Cléo de 5 à 7) (1962)	84
3741	Badlands (1973)	1038
3813	Interiors (1978)	296
3949	Requiem for a Dream (2000)	5778
3994	Unbreakable (2000)	7109
4007	Wall Street (1987)	2215
4144	In the Mood For Love (Fa yeung nin wa) (2000)	916
4298	Rififi (Du rififi chez les hommes) (1955)	337
4308	Moulin Rouge (2001)	5818
4334	Yi Yi (2000)	315
4343	Evolution (2001)	2353
4369	Fast and the Furious, The (2001)	3298
4393	Another Woman (1988)	100
4406	Man Who Shot Liberty Valance, The (1962)	882
4914	Breathless (À bout de souffle) (1960)	707
4995	Beautiful Mind, A (2001)	9575
5056	Every Man for Himself and God Against All (a.k.a. The Enigma of Kaspar Hauser) (a.k.a. The Mystery of Kaspar Hauser) (Jeder für sich und Gott Gegen Alle) (1974)	163
5236	Tale of Springtime, A (Conte de Printemps) (1990)	41
5266	Panic Room (2002)	3412
5272	Time Out (L'Emploi du temps) (2001)	133
5327	In Praise of Love (Éloge de L'amour) (2001)	34
5349	Spider-Man (2002)	10745
5365	Lady and the Duke, The (L'Anglaise et le Duc) (2001)	52
5378	Star Wars: Episode II - Attack of the Clones (2002)	7934
5388	Insomnia (2002)	3245
5494	Earth Trembles, The (La Terra Trema) (1948)	18

5496	Ossessione (1943)	38
5878	Talk to Her (Hable con Ella) (2002)	1604
5902	Adaptation (2002)	4471
5945	About Schmidt (2002)	2768
5954	25th Hour (2002)	2053
6023	Band of Outsiders (Bande à part) (1964)	179
6162	Gerry (2002)	122
6242	Ring (Ringu) (1998)	1301
6258	Les Dames du Bois de Boulogne (1945)	32
6281	Phone Booth (2002)	2503
6291	Lilya 4-Ever (Lilja 4-ever) (2002)	471
6571	Mondays in the Sun (Los Lunes al sol) (2002)	88
6643	Tokyo Story (Tôkyô monogatari) (1953)	242
6669	Ikiru (1952)	482
6711	Lost in Translation (2003)	7096
6757	Demonlover (2002)	56
6987	Cabinet of Dr. Caligari, The (Das Cabinet des Dr. Caligari.) (1920)	608
6993	Hannah and Her Sisters (1986)	974
7022	Battle Royale (Batoru Rowaiaru) (2000)	1768
7043	My Life to Live (Vivre sa vie: Film en douze tableaux) (1962)	112
7056	Public Enemy, The (1931)	206
7074	Navigator, The (1924)	93
7112	Accident (1967)	19
7136	Stolen Kisses (Baisers volés) (1968)	85
7216	High Sierra (1941)	257
7327	Persona (1966)	353
7361	Eternal Sunshine of the Spotless Mind (2004)	8393
7419	After Hours (1985)	566
7748	Pierrot Le Fou (1965)	105
7761	Soft Skin, The (Peau douce, La) (1964)	18
7766	Throne of Blood (Kumonosu jô) (1957)	453
7880	Friday Night (Vendredi Soir) (2002)	11
7897	Ballad of Cable Hogue, The (1970)	73
8125	Sunrise: A Song of Two Humans (1927)	165
8235	Safety Last! (1923)	135
8516	Matter of Life and Death, A (Stairway to Heaven) (1946)	89
8638	Before Sunset (a.k.a. Before Sunrise 2) (2004)	1326
8684	Man Escaped, A (Un condamné à mort s'est échappé ou Le vent souffle où il veut) (1956)	97
8798	Collateral (2004)	3644
8955	Undertow (2004)	125
9011	Portrait of Jennie (1948)	31
25750	Sherlock Jr. (1924)	131
25782	Bodu Saved From Drowning (Bodu sauvé des eaux) (1932)	4
25833	Camille (1936)	49

25941	Letter from an Unknown Woman (1948)	9
26002	Mr. Arkadin (Confidential Report) (1955)	39
26055	Floating Weeds (Ukigusa) (1959)	59
26796	Heart in Winter, A (Un coeur en hiver) (1992)	29
27266	2046 (2004)	472
27664	Brown Bunny, The (2003)	64
27792	Saddest Music in the World, The (2003)	84
30707	Million Dollar Baby (2004)	3710
30747	Goodbye, Dragon Inn (Bu san) (2003)	28
32179	Elevator to the Gallows (Ascenseur pour l'échafaud) (1958)	94
33493	Star Wars: Episode III - Revenge of the Sith (2005)	5193
39369	Detective Story (1951)	7
44555	Lives of Others, The (Das Leben der Anderen) (2006)	1230
44657	Mouchette (1967)	39
44974	Hard Candy (2005)	516
45081	Silent Hill (2006)	454
48516	Departed, The (2006)	4145
52967	Away from Her (2006)	173
53447	Paranoid Park (2007)	98
55269	Darjeeling Limited, The (2007)	682
55272	We Own the Night (2007)	182
57792	Caramel (Sukkar banat) (2007)	32
58559	Dark Knight, The (2008)	2598
59832	Where the Sidewalk Ends (1950)	4
60391	Aleksandra (2007)	2
61967	Autumn Afternoon, An (Sanma no aji) (1962)	8
63768	Tattooed Life (Irezumi ichidai) (1965)	3